

## THE IL-1 GENE CLUSTER AND ASSOCIATED INFLAMMATORY POLYMORPHISMS AND HAPLOTYPES

### Related Applications

5           This application is a Continuation of U.S.S.N. 351,702, filed January 27, 2003, which claims the benefit of U.S. Provisional Application No. 60/351,951, filed January 25, 2002, the contents of both are hereby incorporated in their entirety

### 1. Background of the Invention

10           IL-1 is a primary inflammatory cytokine and has been implicated in mediating both acute and chronic pathological inflammatory diseases. Two functionally similar molecules, IL-1  $\alpha$  and IL-1 $\beta$ , are encoded by separate genes (respectively, *IL1A* and *IL1B*). The third gene of the family (*IL1RN*) encodes IL-1 receptor antagonist (IL-1ra), an anti-inflammatory non-signaling molecule that competes for receptor binding with IL-1 $\alpha$  and IL-1 $\beta$ . Pairwise  
15           comparison of IL-1 $\alpha$ , IL-1 $\beta$  and IL-1ra yields <25% identity in each case, yet X-ray crystallography of IL-1 $\beta$  and IL-1ra reveal closely similar folds (Priestle et al. (1989) PNAS USA 86: 9667-967); Vigers et al. (1994) Biol Chem 269: 12874-12879). Structurally, the proteins consist of a single domain of 12 packed  $\beta$ -sheets known as a beta-trefoil. Since most of the packing interactions feature main chain atoms, it has been argued  
20           that few invariable amino acid are residues required to produce the IL-1 fold, hence extensive diversification of the coding sequences of the genes has been possible. A very similar fold is achieved in soybean trypsin inhibitor without any detectable sequence similarity. All three proteins bind the only functional signaling receptor for IL-1, the type I IL-1 receptor (IL-1R1) (see Sims et al. (1993) PNAS USA 90: 6155-6159).

25           IL-1 has been characterized mainly as the product of stimulated monocytes, macrophages and keratinocytes, but important roles have been suggested for IL-1 released from smooth muscle and endothelial cells (reviewed by Ross (1993) Nature 362: 801-9). Signaling through IL-1R1 involves the cytoplasmic Toll-like domain of the receptor (Heguy et al. (1992).J Biol Chem 267: 2605-2609). Functional IL-1 receptors are widely  
30           distributed in tissues. It is currently believed that IL-1ra differs from IL-1 in failing to activate the interaction between IL-1R1 and the second receptor component, IL-1 receptor accessory protein, IL-1RacP. This is a transmembrane protein that is a distant relative of IL-1R1, having a similar domain structure, but has no intrinsic affinity for IL-1 (Greenfeder

et al. (1995) J Biol Chem 270: 13757-13756; Wesche et al., (1997) J Biol Chem 272: 7727-7731).

The IL-1 gene cluster is on the long arm of chromosome 2 (2q13) and contains at least the genes for IL-1 $\alpha$  (IL-1A), IL-1 $\beta$  (IL-1B), and the IL-1 receptor antagonist (IL-1RN), within a region of 430 Kb (Nicklin, et al. (1994) Genomics, 19: 382-4). The maximum separation of the distal genes *IL1A* and *IL1RN* has been estimated to be 430 kb by pulse field gel electrophoresis of restriction digests of human genomic DNA (Nicklin, et al. (1994) Genomics, 19: 382-4), and the orientation of the three genes has been determined by sequence analysis of physical clones (Nothwang et al. (1997) Genomics 41: 370-378).

IL-18 appears to be the fourth member of the IL-1 structural family (Bazan et al. (1996) Nature 379: 591). It is also a proinflammatory cytokine, but its activity parallels that of IL-1. IL-18 binds to a related receptor (IL-18R1) rather than IL-1R1 (Torigoe et al. (1997) J Biol Chem 272: 25737-25742), which engages a related accessory protein, IL-18RacP, rather than IL-1RacP (Born et al. (1998). The IL-18 gene, *IL18*, resides on chromosome 11 (Nolan et al., (1998) Genomics 51: 161-3).

Certain other proteins that contain IL-1-like elements have been identified from commercial and public cDNA databases (Mulero et al. (1999) Biochem Biophys Res Commun 5: 702-6; Smith et al. (2000) J Biol Chem 275: 1169-1175); Kumar et al., (2000) J Biol Chem 275: 10308-10314; Busfield et al. (2000) Genomics 66: 213-216; Lin et al. (2001) J Biol Chem 276: 20597-20602). One IL-1 like gene was also identified after cDNA selection by hybridization with a YAC clone that incorporated the IL-1 cluster (Barton et al., (2000) Eur J Immunol 30: 3299-3308). This IL-1 gene and its product (i.e. the Interleukin-1-like protein 1 gene/ product) are described in detail in our pending application U.S.S.N. 09/617720, the contents of which are incorporated herein by reference. A uniform nomenclature system for the six new genes has recently been agreed by the investigators involved in the discovery of the genes (see Sims et al. (2001) Trends Immunol 22: 536-537) and will be used herein. Recognizing the four previously known IL-1 family members, the new human genes have been named *IL1F5* (i.e. *IL-1L1*), *IL1F6*, *IL1F7*, *IL1F8*, *IL1F9* and *IL1F10*. Protein products are named in the style, IL-1F7b (which would mean, the second described putative protein product of the *IL1F7* gene). The genes generally appear to be conserved between man and mouse.

In U.S. Patent No. 6,268,142, the contents of which are hereby incorporated by reference in their entirety, we have previously described certain polymorphisms, including

SNPs, associated with IL-1 inflammatory haplotypes and their use in inflammatory disease diagnostics and therapeutics. In U.S.S.N. 09/617720 and U.S.S.N. 09/969,215 [Publication No. US 2002/0182612}}, the contents of which are hereby incorporated in their entirety, we have previously described therapeutics and diagnostics based on the IL-1 B allele 2 (+6912) polymorphism. Still further, in U.S.S.N. 010/300011 (also PCT US 02/37222), the contents of which are also hereby incorporated in their entirety, we describe and characterize functional polymorphisms, including those in an upstream region of the IL-1 B gene, that affect transcription and susceptibility to inflammatory and infectious disease. In addition, in U.S.S.N. 09/617720, the contents of which are hereby incorporated in their entirety, we previously describe the IL-1 like-gene and it's product (i.e. the Interleukin-1-like protein 1 gene/ product, i.e. IL-1F5). Recognizing that the entire IL-1 gene locus is centrally involved in inflammatory disease, we herein provide further detailed IL-1 locus polymorphism, linkage, disease association and functional analysis supporting compositions for detecting genetic identity at the human IL-1 locus and their use for the prediction, diagnosis and therapy of inflammatory disease.

## 2. Summary of the Invention

In general the invention provides compositions and methods for detecting and IL-1 haplotype (e.g. an IL-1 haplotype associated with an increased risk or a decreased risk of developing an inflammatory disease or condition). In preferred embodiments, the IL-1 haplotype is one associated with either an increased risk or a decreased risk of developing a disease or condition, however the invention necessarily encompasses materials and methods for detecting an IL-1 haplotype associated with neither an increased nor a decreased risk for developing a disease or condition (e.g. a "normal" or "wt" genotype).

In preferred embodiments, the invention provides compositions and methods for determining whether a subject has or is predisposed to developing a disease or condition that is associated with an IL-1 inflammatory haplotype by detecting an IL-1 allele associated with an inflammatory disease or disorder or any IL-1 allele in linkage disequilibrium with such an allele - e.g. one or more linked IL-1 alleles as shown in any of Figures 1, 2A, 2B, 7A or 7B. In preferred embodiments, the linked allele has a linkage disequilibrium value ( $D'$ ) with the inflammatory associated allele of at least 0.5 and preferably at least 0.6, 0.7, 0.8 or 0.9.

In another embodiment, the invention provides compositions and methods for determining whether a subject has a decreased risk for developing a disease or condition that is associated with an IL-1 inflammatory haplotype by detecting an IL-1 allele associated with a decreased risk of the inflammatory disease or disorder or any IL-1 allele that is in linkage disequilibrium with such a “protective” allele - e.g. one or more linked IL-1 alleles as shown in any of Figures 1, 2A, 2B, 7A or 7B. In preferred embodiments, the linked allele has a linkage disequilibrium value (D’) with the “protective allele” of at least 0.5 and preferably at least 0.6, 0.7, 0.8 or 0.9. In certain preferred embodiments, the invention includes 4 new IL-1 haplotypes (hap1-4), based on newly identified SNPs. In one preferred embodiment, the invention provides hap1 (IL-1 haplotype pattern 1) an IL-1 pro-inflammatory (consistent with the previously described haplotype: 3322146121) which includes: the IL-1 A(+4845) allele 2 (in 100% LD with IL-1A(-889) allele 2); the IL-1B(+3954) allele 2; and the IL-1B(-511) allele 1. In another embodiment, the invention provides a hap1 haplotype comprising a multiplicity of two or more alleles of a hap1 haplotype pattern as shown in Figure 3A and 3B. In preferred embodiments the hap1 haplotype includes the IL-1 TTC/2-2-1 pattern indicated in Figure 3A and B.

In another embodiment, the invention provides an IL-1 haplotype, hap2, consistent with the previously described haplotype: 4411233212, which includes: the IL-1 A(+4845) allele 1 (in 100% LD with IL-1A(-889) allele 1); the IL-1B(+3954) allele 1 IL-1B(-511) allele 2. In another embodiment, the invention provides a hap2 haplotype comprising a multiplicity of two or more alleles of a hap2 haplotype pattern as shown in Figure 4A and 4B. In preferred embodiments the hap2 haplotype includes the IL-1 GCT/1-1-2 pattern indicated in Figure 4A and 4B.

In yet another embodiment, the invention provides an IL-1 haplotype, hap3, consistent with the previously described (“wild type”) allelic pattern \*\*111\*\*\* which includes: the IL-1 A(+4845) allele 1 (in 100% LD with IL-1A(-889) allele 1); the IL-1B(+3954) allele 1; and the IL-1B(-511) allele 1. In a preferred embodiment, the invention provides a hap3 haplotype comprising a multiplicity of two or more alleles of a hap3 haplotype pattern as shown in Figure 5A and 5B. In preferred embodiments the hap3 haplotype includes the IL-1 hap3 GCC/1-1-1 pattern indicated in Figure 5A and 5B.

In yet another embodiment, the invention provides newly identified SNPs, that are consistent with a new IL-1 haplotype pattern (hap4) comprising: IL-1B(+3954) allele; IL-1B(-511) allele 1; and IL-1 B(-3737) allele 1. In a preferred embodiment, the invention

provides a hap4 haplotype comprising a multiplicity of two or more alleles of a hap 4 haplotype pattern as shown in Figure 6A and 6B. In preferred embodiments the hap 3 haplotype includes the IL-1 hap4 CCC/1-1-1 pattern indicated in Figure 6A and 6B..

It is further an object of the invention to provide methods and compositions relating to the use of sequence information from the IL-1 gene cluster and, in particular, from the novel IL-1-like genes of the IL-1 cluster. It is a further object to integrate this sequence information with genetic data. Accordingly, the invention provides a map of the IL-1 cluster that provides detailed information on the structure and organization of the genes and associated polymorphisms. It is still further an object of the invention to provide methods of predicting and diagnosing a disease or disorders associated with the IL-1 gene cluster. It is further a goal to provide a multiplicity of human IL-1 gene cluster sequence identifiers, comprising one or more nucleic acids for the identification of an IL-1 polymorphism as shown in Figure 4.

### 3. Brief Description of the Figures

Figure 1 represents schematically the linkage disequilibrium of representative SNPs throughout the IL-1 gene cluster locus.

Figure 2 (A and B) shows representative quantitative values for linkage disequilibrium ( $D'$  values appear below the diagonal) and their statistical significance (1-p values appear above the diagonal) of representative SNPs throughout the IL-1 gene cluster.

Figure 3 (A and B) shows the organization of SNPs of IL-1 haplotype pattern 1 (hap 1) (T-T-C = 2\_2\_1).

Figure 4 (A and B) shows the organization of SNPs of IL-1 haplotype pattern 2 (hap 2) (G-C-T = 1\_1\_2).

Figure 5 (A and B) shows the organization of SNPs of IL-1 haplotype pattern 3 (hap 3) (G-C-C) = 1\_1\_1).

Figure 6 (A and B) shows the organization of SNPs of IL-1 haplotype pattern 4 (hap 4) (C-C-C = 1\_1\_1).

Figure 7 (A and B) shows the SNPs that are in strong linkage disequilibrium and not specifically included in the LD table

Figure 8 shows the identity and position of IL-1A gene polymorphisms.

Figure 9 shows the identity and position of IL-1B gene polymorphisms.

Figure 10 (A and B) shows the identity and position of IL-1RNic gene polymorphisms.

Figure 11 shows the identity and position of IL-1RNsec gene polymorphisms.

Figure 12 shows that the difference in cleavage by calpain protease of IL-1 $\alpha$  variant  
5 corresponding to alleles 1 and 2 of IL-1A +4845.

Figure 13 shows the rate of proliferation of fibroblast cells stably transfected with vectors expressing the allele 1 and allele 2 variants of IL-1+4845.

Figure 14 (A and B) shows the genotypes of IL-1A SNP constructs (A) and selected reporter activities in a fibroblast cell line (B).

10 Figure 15 (A, B, C, and D) shows the genotypes of IL-1B SNP constructs (A) and selected re reporter activities in a fibroblast cell line (B); as well as the genotypes of another set of IL-1B constructs in with allele 2 occurring at positions 14 and 15 (C) and selected reporter activities in a fibroblast cell line (D).

Figure 16 (A and B) shows the genotypes of IL-1RN SNP constructs (A) and  
15 selected reporter activities in a fibroblast cell line (B).

Figure 17 shows a map of the IL-1 gene cluster..Scale bars (in kb) are provided above and below the data to aid alignments.

Figure 18 (A-G) shows the alignment of the encoded sequence of the three common exons of the ten known members of the IL-1 family.

20 Figure 19 shows the map positions of select polymorphic markers within the IL-1 gene cluster.

#### **4. Detailed Description of the Invention**

##### **4.1. General**

25 Several homologs of the cytokine interleukin (IL)-1 gene map to the previously identified IL-1 gene cluster, but the public sequencing of the region has been relatively slow. We have therefore constructed a contig of the entire cluster and annotated it. In addition, novel human polymorphic loci in this gene cluster (including SNPs in IL-1A, IL-1B and IL-RN) and associated IL-1 haplotypes have been located and identified as  
30 summarized in Figures 1-11. The features of the invention are further demonstrated in the accompanying detailed description of the invention and examples which.

#### 4.2. Definitions

For convenience, the meaning of certain terms and phrases employed in the specification, examples, and appended claims is provided below.

The term "allele" refers to the different sequence variants found at different polymorphic regions. For example, IL-1RN (VNTR) has at least five different alleles. The sequence variants may be single or multiple base changes, including without limitation insertions, deletions, or substitutions, or may be a variable number of sequence repeats.

The term "allelic pattern" refers to the identity of an allele or alleles at one or more polymorphic regions. For example, an allelic pattern may consist of a single allele at a polymorphic site, as for IL-1RN (VNTR) allele 1, which is an allelic pattern having at least one copy of IL-1RN allele 1 at the VNTR of the IL-1RN gene loci. Alternatively, an allelic pattern may consist of either a homozygous or heterozygous state at a single polymorphic site. For example, IL1-RN (VNTR) allele 2,2 is an allelic pattern in which there are two copies of the second allele at the VNTR marker of IL-1RN that corresponds to the homozygous IL-RN (VNTR) allele 2 state. Alternatively, an allelic pattern may consist of the identity of alleles at more than one polymorphic site.

The term "antibody" as used herein is intended to refer to a binding agent including a whole antibody or a binding fragment thereof which is specifically reactive with an IL-1 polypeptide. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, F(ab).sub.2 fragments can be generated by treating an antibody with pepsin. The resulting F(ab).sub.2 fragment can be treated to reduce disulfide bridges to produce Fab fragments. The antibody of the present invention is further intended to include bispecific, single-chain, and chimeric and humanized molecules having affinity for an IL-1B polypeptide conferred by at least one CDR region of the antibody.

"Biological activity" or "bioactivity" or "activity" or "biological function", which are used interchangeably, for the purposes herein means an effector or antigenic function that is directly or indirectly performed by an IL-1 polypeptide (whether in its native or denatured conformation), or by any subsequence thereof. Biological activities include binding to a target peptide, e.g., an IL-1 receptor. An IL-1 bioactivity can be modulated by directly affecting an IL-1 polypeptide. Alternatively, an IL-1 bioactivity can be modulated by modulating the level of an IL-1 polypeptide, such as by modulating expression of an IL-1 gene.

As used herein the term "bioactive fragment of an IL-1 polypeptide" refers to a fragment of a full-length IL-1 polypeptide, wherein the fragment specifically mimics or antagonizes the activity of a wild-type IL-1 polypeptide. The bioactive fragment preferably is a fragment capable of interacting with an interleukin receptor.

5       The term "an aberrant activity", as applied to an activity of a polypeptide such as IL-1, refers to an activity which differs from the activity of the wild-type or native polypeptide or which differs from the activity of the polypeptide in a healthy subject. An activity of a polypeptide can be aberrant because it is stronger than the activity of its native counterpart. Alternatively, an activity can be aberrant because it is weaker or absent relative to the  
10       activity of its native counterpart. An aberrant activity can also be a change in an activity. For example an aberrant polypeptide can interact with a different target peptide. A cell can have an aberrant IL-1 activity due to overexpression or underexpression of an IL-1 locus gene encoding an IL-1 locus polypeptide.

      "Cells", "host cells" or "recombinant host cells" are terms used interchangeably  
15       herein to refer not only to the particular subject cell, but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact be identical to the parent cell, but are still included within the scope of the term as used herein.

      A "chimera," "mosaic," "chimeric mammal" and the like, refers to a transgenic  
20       mammal with a knock-out or knock-in construct in at least some of its genome-containing cells.

      The terms "control" or "control sample" refer to any sample appropriate to the detection technique employed. The control sample may contain the products of the allele detection technique employed or the material to be tested. Further, the controls may be  
25       positive or negative controls. By way of example, where the allele detection technique is PCR amplification, followed by size fractionation, the control sample may comprise DNA fragments of an appropriate size. Likewise, where the allele detection technique involves detection of a mutated protein, the control sample may comprise a sample of a mutant protein. However, it is preferred that the control sample comprises the material to be tested.  
30       For example, the controls may be a sample of genomic DNA or a cloned portion of the IL-1 gene cluster. However, where the sample to be tested is genomic DNA, the control sample is preferably a highly purified sample of genomic DNA.



The phrase "diseases and conditions associated with IL-1 polymorphisms" refers to a variety of diseases or conditions, the susceptibility to which can be indicated in a subject based on the identification of one or more alleles within the IL-1 complex. Examples include: inflammatory or degenerative disease, including: Systemic Inflammatory Response (SIRS); Alzheimer's Disease (and associated conditions and symptoms including: chronic neuroinflammation, glial activation; increased microglia; neuritic plaque formation; and response to therapy); Amyotrophic Lateral Sclerosis (ALS), arthritis (and associated conditions and symptoms including: acute joint inflammation, antigen-induced arthritis, arthritis associated with chronic lymphocytic thyroiditis, collagen-induced arthritis, juvenile chronic arthritis; juvenile rheumatoid arthritis, osteoarthritis, prognosis and streptococcus-induced arthritis), asthma (and associated conditions and symptoms, including: bronchial asthma; chronic obstructive airway disease; chronic obstructive pulmonary disease, juvenile asthma and occupational asthma); cardiovascular diseases (and associated conditions and symptoms, including atherosclerosis; autoimmune myocarditis, chronic cardiac hypoxia, congestive heart failure, coronary artery disease, cardiomyopathy and cardiac cell dysfunction, including: aortic smooth muscle cell activation; cardiac cell apoptosis; and immunomodulation of cardiac cell function; diabetes and associated conditions and symptoms, including autoimmune diabetes, insulin-dependent (Type 1) diabetes, diabetic periodontitis, diabetic retinopathy, and diabetic nephropathy); gastrointestinal inflammations (and related conditions and symptoms, including celiac disease, associated osteopenia, chronic colitis, Crohn's disease, inflammatory bowel disease and ulcerative colitis); gastric ulcers; hepatic inflammations, cholesterol gallstones and hepatic fibrosis, HIV infection (and associated conditions and symptoms, including degenerative responses, neurodegenerative responses, and HIV associated Hodgkin's Disease), Kawasaki's Syndrome (and associated diseases and conditions, including mucocutaneous lymph node syndrome, cervical lymphadenopathy, coronary artery lesions, edema, fever, increased leukocytes, mild anemia, skin peeling, rash, conjunctiva redness, thrombocytosis; multiple sclerosis, nephropathies (and associated diseases and conditions, including diabetic nephropathy, endstage renal disease, glomerulonephritis, Goodpasture's syndrome, hemodialysis survival and renal ischemic reperfusion injury), neurodegenerative diseases (and associated diseases and conditions, including acute neurodegeneration, induction of IL-1 in aging and neurodegenerative disease, IL-1 induced plasticity of hypothalamic neurons and chronic stress hyperresponsiveness), Ophthalmopathies (and associated

diseases and conditions, including diabetic retinopathy, Graves Ophthalmopathy, and uveitis, osteoporosis (and associated diseases and conditions, including alveolar, femoral, radial, vertebral or wrist bone loss or fracture incidence, postmenopausal bone loss, mass, fracture incidence or rate of bone loss), otitis media (adult or pediatric), pancreatis or

5 pancreatic acinitis, periodontal disease (and associated diseases and conditions, including adult, early onset and diabetic); pulmonary diseases, including chronic lung disease, chronic sinusitis, hyaline membrane disease, hypoxia and pulmonary disease in SIDS; restenosis; rheumatism including rheumatoid arthritis, rheumatic aschoff bodies, rheumatic diseases and rheumatic myocarditis; thyroiditis including chronic lymphocytic thyroiditis; urinary

10 tract infections including chronic prostatitis, chronic pelvic pain syndrome and urolithiasis. Immunological disorders, including autoimmune diseases, such as alopecia aerata, autoimmune myocarditis, Graves' disease, Graves ophthalmopathy, lichen sclerosis, multiple sclerosis, psoriasis, systemic lupus erythematosus, systemic sclerosis, thyroid diseases (e.g. goiter and struma lymphomatosa (Hashimoto's thyroiditis, lymphadenoid

15 goiter), sleep disorders and chronic fatigue syndrome and obesity (non-diabetic or associated with diabetes). Resistance to infectious diseases, such as Leishmaniasis, Leprosy, Lyme Disease, Lyme Carditis, malaria, cerebral malaria, meningitis, tubulointestinal nephritis associated with malaria), which are caused by bacteria, viruses (e.g. cytomegalovirus, encephalitis, Epstein-Barr Virus, Human Immunodeficiency Virus,

20 Influenza Virus) or protozoans (e.g., Plasmodium falciparum, trypanosomes). Response to trauma, including cerebral trauma (including strokes and ischemias, encephalitis, encephalopathies, epilepsy, perinatal brain injury, prolonged febrile seizures, SIDS and subarachnoid hemorrhage), low birth weight (e.g. cerebral palsy), lung injury (acute hemorrhagic lung injury, Goodpasture's syndrome, acute ischemic reperfusion), myocardial

25 dysfunction, caused by occupational and environmental pollutants (e.g. susceptibility to toxic oil syndrome silicosis), radiation trauma, and efficiency of wound healing responses (e.g. burn or thermal wounds, chronic wounds, surgical wounds and spinal cord injuries). Susceptibility to neoplasias, including breast cancer associated osteolytic metastasis, cachexia, colorectal cancer, hyperproliferative diseases, Hodgkin's disease, leukemias,

30 lymphomas, metabolic diseases and tumors, metastases, myelomas, and various cancers (including breast prostate ovarian, colon, lung, etc), anorexia and cachexia. Hormonal regulation including fertility/fecundity, likelihood of a pregnancy, incidence of preterm labor, prenatal and neonatal complications including preterm low birth weight, cerebral

palsy, septicemia, hypothyroxinemia, oxygen dependence, cranial abnormality, early onset menopause. A subject's response to transplant (rejection or acceptance), acute phase response (e.g. febrile response), general inflammatory response, acute respiratory distress response, acute systemic inflammatory response, wound healing, adhesion, immunoinflammatory response, neuroendocrine response, fever development and resistance, acute-phase response, stress response, disease susceptibility, repetitive motion stress, tennis elbow, and pain management and response.

The phrases "disruption of the gene" and "targeted disruption" or any similar phrase refers to the site specific interruption of a native DNA sequence so as to prevent expression of that gene in the cell as compared to the wild-type copy of the gene. The interruption may be caused by deletions, insertions or modifications to the gene, or any combination thereof.

The term "haplotype" as used herein is intended to refer to a set of alleles that are inherited together as a group (are in linkage disequilibrium) at statistically significant levels (p.sub.corr <0.05). As used herein, the phrase "an IL-1 haplotype" refers to a haplotype in the IL-1 loci. An IL-1 inflammatory or proinflammatory haplotype refers to a haplotype that is indicative of increased agonist and/or decreased antagonist activities.

The terms "IL-1 gene cluster" and "IL-1 loci" as used herein include all the nucleic acid at or near the 2q13 region of chromosome 2, including at least the IL-1A, IL-1B and IL-1RN genes and any other linked sequences. (Nicklin et al., Genomics 19:382-84, 1994). The terms "IL-1A", "IL-1B", and "IL-1RN" as used herein refer to the genes coding for IL-1, IL-1, and IL-1 receptor antagonist, respectively. The gene accession number for IL-1A, IL-1B, and IL-1RN are X03833, X04500, and X64532, respectively.

"IL-1 functional mutation" refers to a mutation within the IL-1 gene cluster that results in an altered phenotype (i.e. effects the function of an IL-1 gene or protein).

Examples include: IL-1A(+4845) allele 2, IL-1B (+3954) allele 2, IL-1B (+6912) allele 2 and IL-1RN (+2018) allele 2.

"IL-1X (Z) allele Y" refers to a particular allelic form, designated Y, occurring at an IL-1 locus polymorphic site in gene X, wherein X is IL-1A, B, or RN and positioned at or near nucleotide Z, wherein nucleotide Z is numbered relative to the major transcriptional start site, which is nucleotide +1, of the particular IL-1 gene X. As further used herein, the term "IL-1X allele (Z)" refers to all alleles of an IL-1 polymorphic site in gene X positioned at or near nucleotide Z. For example, the term "IL-1RN (+2018) allele" refers to alternative forms of the IL-1RN gene at marker +2018. "IL-1RN (+2018) allele 1" refers to a form of

the IL-1RN gene which contains a cytosine (C) at position +2018 of the sense strand. Clay et al., Hum. Genet. 97:723-26, 1996. "IL-1RN (+2018) allele 2" refers to a form of the IL-1RN gene which contains a thymine (T) at position +2018 of the plus strand. When a subject has two identical IL-1RN alleles, the subject is said to be homozygous, or to have the homozygous state. When a subject has two different IL-1RN alleles, the subject is said to be heterozygous, or to have the heterozygous state. The term "IL-1RN (+2018) allele 2,2" refers to the homozygous IL-1 RN (+2018) allele 2 state. Conversely, the term "IL-1RN (+2018) allele 1,1" refers to the homozygous IL-1 RN (+2018) allele 1 state. The term "IL-1RN (+2018) allele 1,2" refers to the heterozygous allele 1 and 2 state.

The term "IL-1 phenotype" is meant to refer to any phenotype resulting from an IL-1 gene locus genetic identity- i.e. including increased and decreased predispositions to an inflammatory disease or condition as well as a "normal" (e.g. average or "wild type") associated likelihood of an inflammatory disease or disorder.

"IL-1 related" as used herein is meant to include all genes related to the human IL-1 locus genes on human chromosome 2 (2q 12-14). These include IL-1 genes of the human IL-1 gene cluster located at chromosome 2 (2q 13-14) which include the IL-1A gene which encodes interleukin-1.alpha., the IL-1B gene which encodes interleukin-1.beta., and the IL-1RN (or IL-1ra) gene which encodes the interleukin-1 receptor antagonist. Furthermore these IL-1 related genes include the type I and type II human IL-1 receptor genes located on human chromosome 2 (2q12) and their mouse homologs located on mouse chromosome 1 at position 19.5 cM. Interleukin-1.alpha., interleukin-1.beta., and interleukin-1RN are related in so much as they all bind to IL-1 type I receptors, however only interleukin-1.alpha. and interleukin-1.beta. are agonist ligands which activate IL-1 type I receptors, while interleukin-1RN is a naturally occurring antagonist ligand. Where the term "IL-1" is used in reference to a gene product or polypeptide, it is meant to refer to all gene products encoded by the interleukin-1 locus on human chromosome 2 (2q 12-14) and their corresponding homologs from other species or functional variants thereof. The term IL-1 thus includes secreted polypeptides which promote an inflammatory response, such as IL-1a and IL-1.beta., as well as a secreted polypeptide which antagonize inflammatory responses, such as IL-1 receptor antagonist and the IL-1 type II (decoy) receptor.

An "IL-1 receptor" or "IL-1R" refers to various cell membrane bound protein receptors capable of binding to and/or transducing a signal from an IL-1 locus-encoded ligand. The term applies to any of the proteins which are capable of binding interleukin-1

(IL-1) molecules and, in their native configuration as mammalian plasma membrane proteins, presumably play a role in transducing the signal provided by IL-1 to a cell. As used herein, the term includes analogs of native proteins with IL-1-binding or signal transducing activity. Examples include the human and murine IL-1 receptors described in U.S. Pat. No. 4,968,607. The term "IL-1 nucleic acid" refers to a nucleic acid encoding an IL-1 protein.

An "IL-1 polypeptide" and "IL-1 protein" are intended to encompass polypeptides comprising the amino acid sequence encoded by the IL-1 genomic DNA sequences shown in FIGS. 1, 2, and 3, or fragments thereof, and homologs thereof and include agonist and antagonist polypeptides.

"Increased risk" refers to a statistically higher frequency of occurrence of the disease or condition in an individual carrying a particular polymorphic allele in comparison to the frequency of occurrence of the disease or condition in a member of a population that does not carry the particular polymorphic allele.

"Decreased risk" refers to a statistically lower frequency of occurrence of the disease or condition in an individual carrying a particular polymorphic allele in comparison to the frequency of occurrence of the disease or condition in a member of a population that does not carry the particular polymorphic allele or in the population as a whole.

The term "interact" as used herein is meant to include detectable relationships or associations (e.g. biochemical interactions) between molecules, such as interactions between protein-protein, protein-nucleic acid, nucleic acid-nucleic acid and protein-small molecule or nucleic acid-small molecule in nature.

The term "isolated" as used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs, or RNAs, respectively, that are present in the natural source of the macromolecule. For example, an isolated nucleic acid encoding one of the subject IL-1 polypeptides preferably includes no more than 10 kilobases (kb) of nucleic acid sequence which naturally immediately flanks the IL-1 gene in genomic DNA, more preferably no more than 5 kb of such naturally occurring flanking sequences, and most preferably less than 1.5 kb of such naturally occurring flanking sequence. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid

fragments which are not naturally occurring as fragments and would not be found in the natural state. The term "isolated" is also used herein to refer to polypeptides which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides.

5           A "knock-in" transgenic animal refers to an animal that has had a modified gene introduced into its genome and the modified gene can be of exogenous or endogenous origin.

          A "knock-out" transgenic animal refers to an animal in which there is partial or complete suppression of the expression of an endogenous gene (e.g, based on deletion of at least a portion of the gene, replacement of at least a portion of the gene with a second  
10           sequence, introduction of stop codons, the mutation of bases encoding critical amino acids, or the removal of an intron junction, etc.).

          A "knock-out construct" refers to a nucleic acid sequence that can be used to decrease or suppress expression of a protein encoded by endogenous DNA sequences in a  
15           cell. In a simple example, the knock-out construct is comprised of a gene, such as the IL-1RN gene, with a deletion in a critical portion of the gene, so that active protein cannot be expressed therefrom. Alternatively, a number of termination codons can be added to the native gene to cause early termination of the protein or an intron junction can be inactivated. In a typical knock-out construct, some portion of the gene is replaced with a  
20           selectable marker (such as the neo gene) so that the gene can be represented as follows: IL-1RN 5'/neo/IL-1RN 3', where IL-1RN5' and IL-1RN 3', refer to genomic or cDNA sequences which are, respectively, upstream and downstream relative to a portion of the IL-1RN gene and where neo refers to a neomycin resistance gene. In another knock-out construct, a second selectable marker is added in a flanking position so that the gene can be  
25           represented as: IL-1RN/neo/IL-1RN/TK, where TK is a thymidine kinase gene which can be added to either the IL-1RN5' or the IL-1RN3' sequence of the preceding construct and which further can be selected against (i.e. is a negative selectable marker) in appropriate media. This two-marker construct allows the selection of homologous recombination events, which removes the flanking TK marker, from non-homologous recombination  
30           events which typically retain the TK sequences. The gene deletion and/or replacement can be from the exons, introns, especially intron junctions, and/or the regulatory regions such as promoters.

"Linkage disequilibrium" refers to co-inheritance of two alleles at frequencies greater than would be expected from the separate frequencies of occurrence of each allele in a given control population. The expected frequency of occurrence of two alleles that are inherited independently is the frequency of the first allele multiplied by the frequency of the second allele. Alleles that co-occur at expected frequencies are said to be in "linkage disequilibrium". The cause of linkage disequilibrium is often unclear. It can be due to selection for certain allele combinations or to recent admixture of genetically heterogeneous populations. In addition, in the case of markers that are very tightly linked to a disease gene, an association of an allele (or group of linked alleles) with the disease gene is expected if the disease mutation occurred in the recent past, so that sufficient time has not elapsed for equilibrium to be achieved through recombination events in the specific chromosomal region. When referring to allelic patterns that are comprised of more than one allele, a first allelic pattern is in linkage disequilibrium with a second allelic pattern if all the alleles that comprise the first allelic pattern are in linkage disequilibrium with at least one of the alleles of the second allelic pattern. An example of linkage disequilibrium is that which occurs between the alleles at the IL-1RN (+2018) and IL-1RN (VNTR) polymorphic sites. The two alleles at IL-1RN (+2018) are 100% in linkage disequilibrium with the two most frequent alleles of IL-1RN (VNTR), which are allele 1 and allele 2.

The term "marker" refers to a sequence in the genome that is known to vary among individuals. For example, the IL-1RN gene has a marker that consists of a variable number of tandem repeats (VNTR).

A "mutated gene" or "mutation" or "functional mutation" refers to an allelic form of a gene, which is capable of altering the phenotype of a subject having the mutated gene relative to a subject which does not have the mutated gene. The altered phenotype caused by a mutation can be corrected or compensated for by certain agents. If a subject must be homozygous for this mutation to have an altered phenotype, the mutation is said to be recessive. If one copy of the mutated gene is sufficient to alter the phenotype of the subject, the mutation is said to be dominant. If a subject has one copy of the mutated gene and has a phenotype that is intermediate between that of a homozygous and that of a heterozygous subject (for that gene), the mutation is said to be co-dominant.

A "non-human animal" of the invention includes mammals such as rodents, non-human primates, sheep, dogs, cows, goats, etc. amphibians, such as members of the *Xenopus* genus, and transgenic avians (e.g. chickens, birds, etc.). The term "chimeric

animal" is used herein to refer to animals in which the recombinant gene is found, or in which the recombinant gene is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that one of the recombinant IL-1 genes is present and/or expressed or disrupted in some tissues but not others. The term "non-human mammal" refers to any member of the class Mammalia, except for humans.

As used herein, the term "nucleic acid" refers to polynucleotides or oligonucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs (e.g. peptide nucleic acids) and as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

The term "nutraceutical", as used herein includes the FDA definitions of foods and dietary supplements that may be of value in treating a disease or disorder- particularly a disease or disorder associated with an inflammatory disease. Accordingly, "nutraceuticals" include nutritional ingredients that can be used to achieve health benefits. These ingredients may be in "foods" ---i.e. "functional foods" or in dietary supplements. In October 1994, the Dietary Supplement Health and Education Act ("DSHEA") was signed into law. DSHEA acknowledges that millions of consumers believe that dietary supplements may provide health benefits. Congress's intent in passing it was to strike a balance between consumer access to dietary supplements and FDA's authority to act against supplements that present safety problems or bear false or misleading labeling. DSHEA creates a new regulatory framework for the safety and labeling of dietary supplements. The FDA is committed to enforcing DSHEA in a manner that effectuates DSHEA. Accordingly, "nutraceuticals," as used herein, includes

dietary supplements known in the art (e.g. vitamins, minerals, herbs and other supplements) which are ingested and are intended to supplement the diet and include a "dietary ingredient." Dietary ingredients may include vitamins, minerals, herbs or other botanicals, amino acids, and dietary substances such as enzymes. Dietary ingredients also can be metabolites, constituents, extracts, concentrates, or combinations of these ingredients. Nutraceutical supplements come in forms including tablets, capsules, liquids, and bars.

The term "polymorphism" refers to the coexistence of more than one form of a gene or portion (e.g., allelic variant) thereof. A portion of a gene of which there are at least two different forms, i.e., two different nucleotide sequences, is referred to as a "polymorphic



region of a gene". A specific genetic sequence at a polymorphic region of a gene is an allele. A polymorphic region can be a single nucleotide, the identity of which differs in different alleles. A polymorphic region can also be several nucleotides long.

The term "propensity to disease," also "predisposition" or "susceptibility" to disease  
5 or any similar phrase, means that certain alleles are hereby discovered to be associated with or predictive of a subject's incidence of developing a particular disease (e.g. a vascular disease). The alleles are thus over-represented in frequency in individuals with disease as compared to healthy individuals. Thus, these alleles can be used to predict disease even in pre-symptomatic or pre-diseased individuals.

10 "Small molecule" as used herein, is meant to refer to a composition, which has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be nucleic acids, peptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules.

As used herein, the term "specifically hybridizes" or "specifically detects" refers to  
15 the ability of a nucleic acid molecule to hybridize to at least approximately 6 consecutive nucleotides of a sample nucleic acid.

"Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, which induce or control transcription of protein coding sequences with which  
20 they are operably linked.

As used herein, the term "transgene" means a nucleic acid sequence (encoding, e.g., one of the IL-1 polypeptides, or an antisense transcript thereto) which has been introduced into a cell. A transgene could be partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous  
25 gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can also be present in a cell in the form of an episome. A transgene can include one or more transcriptional regulatory  
30 sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid.

A "transgenic animal" refers to any animal, preferably a non-human mammal, bird or an amphibian, in which one or more of the cells of the animal contain heterologous

nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation  
 5 does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic animals described herein, the transgene causes cells to express a recombinant form of one of an IL-1 polypeptide, e.g. either agonistic or antagonistic forms. However, transgenic  
 10 animals in which the recombinant gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. Moreover, "transgenic animal" also includes those recombinant animals in which gene disruption of one or more genes is caused by human intervention, including both recombination and antisense techniques. The term is intended to include all progeny generations. Thus, the founder  
 15 animal and all F1, F2, F3, and so on, progeny thereof are included.

The term "treating" as used herein is intended to encompass curing as well as ameliorating at least one symptom of a condition or disease.

The term "vector" refers to a nucleic acid molecule, which is capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome,  
 20 i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally  
 25 to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

30 The term "wild-type allele" refers to an allele of a gene which, when present in two copies in a subject results in a wild-type phenotype. There can be several different wild-type alleles of a specific gene, since certain nucleotide changes in a gene may not affect the phenotype of a subject having two copies of the gene with the nucleotide changes.

#### 4.3 Detection of Alleles

Many methods are available for detecting specific alleles at human polymorphic loci. The preferred method for detecting a specific polymorphic allele will depend, in part, upon the molecular nature of the polymorphism. For example, the various allelic forms of the polymorphic locus may differ by a single base-pair of the DNA. Such single nucleotide polymorphisms (or SNPs) are major contributors to genetic variation, comprising some 80% of all known polymorphisms, and their density in the human genome is estimated to be on average 1 per 1,000 base pairs. SNPs are most frequently biallelic--occurring in only two different forms (although up to four different forms of an SNP, corresponding to the four different nucleotide bases occurring in DNA, are theoretically possible). Nevertheless, SNPs are mutationally more stable than other polymorphisms, making them suitable for association studies in which linkage disequilibrium between markers and an unknown variant is used to map disease-causing mutations. In addition, because SNPs typically have only two alleles, they can be genotyped by a simple plus/minus assay rather than a length measurement, making them more amenable to automation.

A variety of methods are available for detecting the presence of a particular single nucleotide polymorphic allele in an individual. Advancements in this field have provided accurate, easy, and inexpensive large-scale SNP genotyping. Most recently, for example, several new techniques have been described including dynamic allele-specific hybridization (DASH), microplate array diagonal gel electrophoresis (MADGE), pyrosequencing, oligonucleotide-specific ligation, the TaqMan system as well as various DNA "chip" technologies such as the Affymetrix SNP chips. These methods require amplification of the target genetic region, typically by PCR. Still other newly developed methods, based on the generation of small signal molecules by invasive cleavage followed by mass spectrometry or immobilized padlock probes and rolling-circle amplification, might eventually eliminate the need for PCR. Several of the methods known in the art for detecting specific single nucleotide polymorphisms are summarized below. The method of the present invention is understood to include all available methods.

Several methods have been developed to facilitate analysis of single nucleotide polymorphisms. In one embodiment, the single base polymorphism can be detected by using a specialized exonuclease-resistant nucleotide, as disclosed, e.g., in Mundy, C. R. (U.S. Pat. No. 4,656,127). According to the method, a primer complementary to the allelic sequence immediately 3' to the polymorphic site is permitted to hybridize to a target

molecule obtained from a particular animal or human. If the polymorphic site on the target molecule contains a nucleotide that is complementary to the particular exonuclease-resistant nucleotide derivative present, then that derivative will be incorporated onto the end of the hybridized primer. Such incorporation renders the primer resistant to exonuclease, and  
5 thereby permits its detection. Since the identity of the exonuclease-resistant derivative of the sample is known, a finding that the primer has become resistant to exonucleases reveals that the nucleotide present in the polymorphic site of the target molecule was complementary to that of the nucleotide derivative used in the reaction. This method has the advantage that it does not require the determination of large amounts of extraneous  
10 sequence data.

In another embodiment of the invention, a solution-based method is used for determining the identity of the nucleotide of a polymorphic site. Cohen, D. et al. (French Patent 2,650,840; PCT Appln. No. W091/02087). As in the Mundy method of U.S. Pat. No. 4,656,127, a primer is employed that is complementary to allelic sequences immediately 3'  
15 to a polymorphic site. The method determines the identity of the nucleotide of that site using labeled dideoxynucleotide derivatives, which, if complementary to the nucleotide of the polymorphic site will become incorporated onto the terminus of the primer.

An alternative method, known as Genetic Bit Analysis or GBA.TM. is described by Goelet, P. et al. (PCT Appln. No. 92/15712). The method of Goelet, P. et al. uses mixtures  
20 of labeled terminators and a primer that is complementary to the sequence 3' to a polymorphic site. The labeled terminator that is incorporated is thus determined by, and complementary to, the nucleotide present in the polymorphic site of the target molecule being evaluated. In contrast to the method of Cohen et al. (French Patent 2,650,840; PCT Appln. No. W091/02087) the method of Goelet, P. et al. is preferably a heterogeneous  
25 phase assay, in which the primer or the target molecule is immobilized to a solid phase.

Recently, several primer-guided nucleotide incorporation procedures for assaying polymorphic sites in DNA have been described (Komher, J. S. et al., Nucl. Acids. Res. 17:7779-7784 (1989); Sokolov, B. P., Nucl. Acids Res. 18:3671 (1990); Syvanen, A.-C., et al., Genomics 8:684-692 (1990); Kuppuswamy, M. N. et al., Proc. Natl. Acad. Sci. (U.S.A.)  
30 88:1143-1147 (1991); Prezant, T. R. et al., Hum. Mutat. 1:159-164 (1992); Ugozzoli, L. et al., GATA 9:107-112 (1992); Nyren, P. et al., Anal. Biochem. 208:171-175 (1993)). These methods differ from GBA.TM. in that they all rely on the incorporation of labeled deoxynucleotides to discriminate between bases at a polymorphic site. In such a format,

since the signal is proportional to the number of deoxynucleotides incorporated, polymorphisms that occur in runs of the same nucleotide can result in signals that are proportional to the length of the run (Syvanen, A. -C., et al., *Amer. J. Hum. Genet.* 52:46-59 (1993)).

5 For mutations that produce premature termination of protein translation, the protein truncation test (PTT) offers an efficient diagnostic approach (Roest, et. al., (1993) *Hum. Mol Genet.* 2:1719-21; van der Lijdt, et. al., (1994) *Genomics* 20:1-4). For PTT, RNA is initially isolated from available tissue and reverse-transcribed, and the segment of interest is amplified by PCR. The products of reverse transcription PCR are then used as a template  
10 for nested PCR amplification with a primer that contains an RNA polymerase promoter and a sequence for initiating eukaryotic translation. After amplification of the region of interest, the unique motifs incorporated into the primer permit sequential in vitro transcription and translation of the PCR products. Upon sodium dodecyl sulfate-polyacrylamide gel electrophoresis of translation products, the appearance of truncated polypeptides signals the  
15 presence of a mutation that causes premature termination of translation. In a variation of this technique, DNA (as opposed to RNA) is used as a PCR template when the target region of interest is derived from a single exon.

Any cell type or tissue may be utilized to obtain nucleic acid samples for use in the diagnostics described herein. In a preferred embodiment, the DNA sample is obtained from  
20 a bodily fluid, e.g. blood, obtained by known techniques (e.g. venipuncture) or saliva. Alternatively, nucleic acid tests can be performed on dry samples (e.g. hair or skin). When using RNA or protein, the cells or tissues that may be utilized must express an IL-1 gene.

Diagnostic procedures may also be performed in situ directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no  
25 nucleic acid purification is necessary. Nucleic acid reagents may be used as probes and/or primers for such in situ procedures (see, for example, Nuovo, G. J., 1992, *PCR in situ hybridization: protocols and applications*, Raven Press, N.Y.).

In addition to methods which focus primarily on the detection of one nucleic acid sequence, profiles may also be assessed in such detection schemes. Fingerprint profiles may  
30 be generated, for example, by utilizing a differential display procedure, Northern analysis and/or RT-PCR.

A preferred detection method is allele specific hybridization using probes overlapping a region of at least one allele of an IL-1 proinflammatory haplotype and having

about 5, 10, 20, 25, or 30 nucleotides around the mutation or polymorphic region. In a preferred embodiment of the invention, several probes capable of hybridizing specifically to other allelic variants involved in a restenosis are attached to a solid phase support, e.g., a "chip" (which can hold up to about 250,000 oligonucleotides). Oligonucleotides can be bound to a solid support by a variety of processes, including lithography. Mutation detection analysis using these chips comprising oligonucleotides, also termed "DNA probe arrays" is described e.g., in Cronin et al. (1996) Human Mutation 7:244. In one embodiment, a chip comprises all the allelic variants of at least one polymorphic region of a gene. The solid phase support is then contacted with a test nucleic acid and hybridization to the specific probes is detected. Accordingly, the identity of numerous allelic variants of one or more genes can be identified in a simple hybridization experiment.

These techniques may also comprise the step of amplifying the nucleic acid before analysis. Amplification techniques are known to those of skill in the art and include, but are not limited to cloning, polymerase chain reaction (PCR), polymerase chain reaction of specific alleles (ASA), ligase chain reaction (LCR), nested polymerase chain reaction, self sustained sequence replication (Guatelli, J. C. et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh, D. Y. et al., 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177), and Q- Beta Replicase (Lizardi, P. M. et al., 1988, Bio/Technology 6:1197).

Amplification products may be assayed in a variety of ways, including size analysis, restriction digestion followed by size analysis, detecting specific tagged oligonucleotide primers in the reaction products, allele-specific oligonucleotide (ASO) hybridization, allele specific 5' exonuclease detection, sequencing, hybridization, and the like.

PCR based detection means can include multiplex amplification of a plurality of markers simultaneously. For example, it is well known in the art to select PCR primers to generate PCR products that do not overlap in size and can be analyzed simultaneously. Alternatively, it is possible to amplify different markers with primers that are differentially labeled and thus can each be differentially detected. Of course, hybridization based detection means allow the differential detection of multiple PCR products in a sample.

Other techniques are known in the art to allow multiplex analyses of a plurality of markers.

In a merely illustrative embodiment, the method includes the steps of (i) collecting a sample of cells from a patient, (ii) isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, (iii) contacting the nucleic acid sample with one or more

primers which specifically hybridize 5' and 3' to at least one allele of an IL-1 proinflammatory haplotype under conditions such that hybridization and amplification of the allele occurs, and (iv) detecting the amplification product. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in  
5 very low numbers.

In a preferred embodiment of the subject assay, the allele of an IL-1 proinflammatory haplotype is identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel  
10 electrophoresis.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the allele. Exemplary sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) Proc. Natl Acad Sci USA 74:560) or Sanger (Sanger et al (1977) Proc. Nat. Acad. Sci USA 74:5463). It is also  
15 contemplated that any of a variety of automated sequencing procedures may be utilized when performing the subject assays (see, for example Biotechniques (1995) 19:448), including sequencing by mass spectrometry (see, for example PCT publication WO 94/16101; Cohen et al. (1996) Adv Chromatogr 36:127-162; and Griffin et al. (1993) Appl Biochem Biotechnol 38:147-159). It will be evident to one of skill in the art that, for certain  
20 embodiments, the occurrence of only one, two or three of the nucleic acid bases need be determined in the sequencing reaction. For instance, A-track or the like, e.g., where only one nucleic acid is detected, can be carried out.

In a further embodiment, protection from cleavage agents (such as a nuclease, hydroxylamine or osmium tetroxide and with piperidine) can be used to detect mismatched  
25 bases in RNA/RNA or RNA/DNA or DNA/DNA heteroduplexes (Myers, et al. (1985) Science 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type allele with the sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to base pair  
30 mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digest the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in

order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton et al (1988) Proc. Natl Acad Sci USA 85:4397; and Saleeba et al (1992) Methods Enzymol. 217:286-295. In a preferred embodiment, the  
5 control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes). For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T  
10 mismatches (Hsu et al. (1994) Carcinogenesis 15:1657-1662). According to an exemplary embodiment, a probe based on an allele of an IL-1 locus haplotype is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify an IL-1 locus allele. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) Proc Natl. Acad. Sci USA 86:2766, see also Cotton (1993) Mutat Res 285:125-144; and Hayashi (1992) Genet Anal Tech Appl 9:73-79). Single-  
20 stranded DNA fragments of sample and control IL-1 locus alleles are denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather  
25 than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) Trends Genet 7:5).

In yet another embodiment, the movement of alleles in polyacrylamide gels  
30 containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al. (1985) Nature 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a



further embodiment, a temperature gradient is used in place of a denaturing agent gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) Biophys Chem 265:12753).

5 Examples of other techniques for detecting alleles include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation or nucleotide difference (e.g., in allelic variants) is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) Nature 324:163); Saiki et al (1989) Proc. Natl Acad. Sci  
10 USA 86:6230). Such allele specific oligonucleotide hybridization techniques may be used to test one mutation or polymorphic region per reaction when oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations or polymorphic regions when the oligonucleotides are attached to the hybridizing membrane and hybridized with labelled target DNA.

15 Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation or polymorphic region of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al (1989) Nucleic Acids Res. 17:2437-2448) or at the extreme 3'  
20 end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) Tibtech 11:238. In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al (1992) Mol. Cell Probes 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification  
25 (Barany (1991) Proc. Natl. Acad. Sci USA 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

30 In another embodiment, identification of the allelic variant is carried out using an oligonucleotide ligation assay (OLA), as described, e.g., in U.S. Pat. No. 4,998,617 and in Landegren, U. et al. ((1988) Science 241:1077-1080). The OLA protocol uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target. One of the oligonucleotides is linked to a separation marker, e.g.,

biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate. Ligation then permits the labeled oligonucleotide to be recovered using avidin, or another biotin ligand. Nickerson, D. A. et al. have described a  
 5 nucleic acid detection assay that combines attributes of PCR and OLA (Nickerson, D. A. et al. (1990) Proc. Natl. Acad. Sci. USA 87:8923-27). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Several techniques based on this OLA method have been developed and can be used to detect alleles of an IL-1 locus haplotype. For example, U.S. Pat. No. 5,593,826 discloses  
 10 an OLA using an oligonucleotide having 3'-amino group and a 5'-phosphorylated oligonucleotide to form a conjugate having a phosphoramidate linkage. In another variation of OLA described in Tobe et al. ((1996) Nucleic Acids Res 24: 3728), OLA combined with PCR permits typing of two alleles in a single microtiter well. By marking each of the allele-specific primers with a unique hapten, i.e. digoxigenin and fluorescein, each OLA reaction  
 15 can be detected by using hapten specific antibodies that are labeled with different enzyme reporters, alkaline phosphatase or horseradish peroxidase. This system permits the detection of the two alleles using a high throughput format that leads to the production of two different colors.

Another embodiment of the invention is directed to kits for detecting a  
 20 predisposition for developing a restenosis. This kit may contain one or more oligonucleotides, including 5' and 3' oligonucleotides that hybridize 5' and 3' to at least one allele of an IL-1 locus haplotype. PCR amplification oligonucleotides should hybridize between 25 and 2500 base pairs apart, preferably between about 100 and about 500 bases apart, in order to produce a PCR product of convenient size for subsequent analysis.

Particularly preferred primers included nucleotide sequences described in Figures 8-  
 25 11. The design of additional oligonucleotides for use in the amplification and detection of IL-1 polymorphic alleles by the method of the invention is facilitated by the availability of both updated sequence information from human chromosome 2q13--which contains the human IL-1 locus, and updated human polymorphism information available for this locus.  
 30 For example, the DNA sequence for the IL-1A, IL-1B and IL-1RN is shown in FIGS. 1 (GenBank Accession No. X03833), 2 (GenBank Accession No. X04500) and 3 (GenBank Accession No. X64532) respectively. Suitable primers for the detection of a human polymorphism in these genes can be readily designed using this sequence information and

standard techniques known in the art for the design and optimization of primers sequences. Optimal design of such primer sequences can be achieved, for example, by the use of commercially available primer selection programs such as Primer 2.1, Primer 3 or GeneFisher (See also, Nicklin M. H. J., Weith A. Duff G. W., "A Physical Map of the  
 5 Region Encompassing the Human Interleukin-1.alpha., interleukin-1.beta., and Interleukin-1 Receptor Antagonist Genes" Genomics 19: 382 (1995); Nothwang H. G., et al. "Molecular Cloning of the Interleukin-1 gene Cluster: Construction of an Integrated YAC/PAC Contig and a partial transcriptional Map in the Region of Chromosome 2q13" Genomics 41:370 (1997); Clark, et al. (1986) Nucl. Acids. Res., 14:7897-7914 [published  
 10 erratum appears in Nucleic Acids Res., 15:868 (1987) and the Genome Database (GDB) project at the URL <http://www.gdb.org>).

For use in a kit, oligonucleotides may be any of a variety of natural and/or synthetic compositions such as synthetic oligonucleotides, restriction fragments, cDNAs, synthetic peptide nucleic acids (PNAs), and the like. The assay kit and method may also employ  
 15 labeled oligonucleotides to allow ease of identification in the assays. Examples of labels which may be employed include radio-labels, enzymes, fluorescent compounds, streptavidin, avidin, biotin, magnetic moieties, metal binding moieties, antigen or antibody moieties, and the like.

The kit may, optionally, also include DNA sampling means. DNA sampling means  
 20 are well known to one of skill in the art and can include, but not be limited to substrates, such as filter papers, the AmpliCard.TM. (University of Sheffield, Sheffield, England S10 2JF; Tarlow, J W, et al., J of Invest. Dermatol. 103:387-389 (1994)) and the like; DNA purification reagents such as Nucleon.TM. kits, lysis buffers, proteinase solutions and the like; PCR reagents, such as 10x reaction buffers, thermostable polymerase, dNTPs, and the  
 25 like; and allele detection means such as the HinfI restriction enzyme, allele specific oligonucleotides, degenerate oligonucleotide primers for nested PCR from dried blood.

#### 4.4. Pharmacogenomics

Knowledge of the particular alleles associated with a susceptibility to developing a particular disease or condition, alone or in conjunction with information on other genetic  
 30 defects contributing to the particular disease or condition allows a customization of the prevention or treatment in accordance with the individual's genetic profile, the goal of "pharmacogenomics". Thus, comparison of an individual's IL-1 profile to the population profile for a vascular disorder, permits the selection or design of drugs or other therapeutic

regimens that are expected to be safe and efficacious for a particular patient or patient population (i.e., a group of patients having the same genetic alteration).

In addition, the ability to target populations expected to show the highest clinical benefit, based on genetic profile can enable: 1) the repositioning of already marketed drugs; 2) the rescue of drug candidates whose clinical development has been discontinued as a result of safety or efficacy limitations, which are patient subgroup-specific; and 3) an accelerated and less costly development for candidate therapeutics and more optimal drug labeling (e.g. since measuring the effect of various doses of an agent on the causative mutation is useful for optimizing effective dose).

The treatment of an individual with a particular therapeutic can be monitored by determining protein (e.g. IL-1.alpha., IL-1.beta., or IL-1Ra), mRNA and/or transcriptional level. Depending on the level detected, the therapeutic regimen can then be maintained or adjusted (increased or decreased in dose). In a preferred embodiment, the effectiveness of treating a subject with an agent comprises the steps of: (i) obtaining a preadministration sample from a subject prior to administration of the agent; (ii) detecting the level or amount of a protein, mRNA or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the protein, mRNA or genomic DNA in the post-administration sample; (v) comparing the level of expression or activity of the protein, mRNA or genomic DNA in the preadministration sample with the corresponding protein, mRNA or genomic DNA in the postadministration sample, respectively; and (vi) altering the administration of the agent to the subject accordingly.

Cells of a subject may also be obtained before and after administration of a therapeutic to detect the level of expression of genes other than an IL-1 gene to verify that the therapeutic does not increase or decrease the expression of genes which could be deleterious. This can be done, e.g., by using the method of transcriptional profiling. Thus, mRNA from cells exposed in vivo to a therapeutic and mRNA from the same type of cells that were not exposed to the therapeutic could be reverse transcribed and hybridized to a chip containing DNA from numerous genes, to thereby compare the expression of genes in cells treated and not treated with the therapeutic.

#### 4.5. Therapeutics For Diseases and Conditions Associated with IL-1 Polymorphisms

Therapeutic for diseases or conditions associated with an IL-1 polymorphism or haplotype refers to any agent or therapeutic regimen (including pharmaceuticals,

nutraceuticals and surgical means) that prevents or postpones the development of or alleviates the symptoms of the particular disease or condition in the subject. The therapeutic can be a polypeptide, peptidomimetic, nucleic acid or other inorganic or organic molecule, preferably a "small molecule" including vitamins, minerals and other nutrients. Preferably the therapeutic can modulate at least one activity of an IL-1 polypeptide, e.g., interaction with a receptor, by mimicking or potentiating (agonizing) or inhibiting (antagonizing) the effects of a naturally-occurring polypeptide. An agonist can be a wild-type protein or derivative thereof having at least one bioactivity of the wild-type, e.g., receptor binding activity. An agonist can also be a compound that upregulates expression of a gene or which increases at least one bioactivity of a protein. An agonist can also be a compound which increases the interaction of a polypeptide with another molecule, e.g., a receptor. An antagonist can be a compound which inhibits or decreases the interaction between a protein and another molecule, e.g., a receptor or an agent that blocks signal transduction or post-translation processing (e.g., IL-1 converting enzyme (ICE) inhibitor). Accordingly, a preferred antagonist is a compound which inhibits or decreases binding to a receptor and thereby blocks subsequent activation of the receptor. An antagonist can also be a compound that downregulates expression of a gene or which reduces the amount of a protein present. The antagonist can be a dominant negative form of a polypeptide, e.g., a form of a polypeptide which is capable of interacting with a target peptide, e.g., a receptor, but which does not promote the activation of the receptor. The antagonist can also be a nucleic acid encoding a dominant negative form of a polypeptide, an antisense nucleic acid, or a ribozyme capable of interacting specifically with an RNA. Yet other antagonists are molecules which bind to a polypeptide and inhibit its action. Such molecules include peptides, e.g., forms of target peptides which do not have biological activity, and which inhibit binding to receptors. Thus, such peptides will bind to the active site of a protein and prevent it from interacting with target peptides. Yet other antagonists include antibodies that specifically interact with an epitope of a molecule, such that binding interferes with the biological function of the polypeptide. In yet another preferred embodiment, the antagonist is a small molecule, such as a molecule capable of inhibiting the interaction between a polypeptide and a target receptor. Alternatively, the small molecule can function as an antagonist by interacting with sites other than the receptor binding site.

Modulators of IL-1 (e.g. IL-1.alpha., IL-1.beta. or IL-1 receptor antagonist) or a protein encoded by a gene that is in linkage disequilibrium with an IL-1 gene can comprise

any type of compound, including a protein, peptide, peptidomimetic, small molecule, or nucleic acid. Preferred agonists include nucleic acids (e.g. encoding an IL-1 protein or a gene that is up- or down-regulated by an IL-1 protein), proteins (e.g. IL-1 proteins or a protein that is up- or down-regulated thereby) or a small molecule (e.g. that regulates expression or binding of an IL-1 protein). Preferred antagonists, which can be identified, for example, using the assays described herein, include nucleic acids (e.g. single (antisense) or double stranded (triplex) DNA or PNA and ribozymes), protein (e.g. antibodies) and small molecules that act to suppress or inhibit IL-1 transcription and/or protein activity.

#### 4.6. Effective Dose and Formulations and Use

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining The LD.sub.50 (the dose lethal to 50% of the population) and the ED.sub.50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD.sub.50 /ED.sub.50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissues in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED.sub.50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC.sub.50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

Compositions for use in accordance with the present invention may be formulated in a conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvates may be

formulated for administration by, for example, injection, inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral or rectal administration.

For such therapy, the compounds of the invention can be formulated for a variety of loads of administration, including systemic and topical or localized administration.

5 Techniques and formulations generally may be found in Remington's Pharmaceutical Sciences, Meade Publishing Co., Easton, Pa. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For injection, the compounds of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In  
10 addition, the compounds may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

For oral administration, the compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinised maize starch, polyvinylpyrrolidone or  
15 hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulfate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions,  
20 or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., ationd oil, oily esters, ethyl alcohol or fractionated  
25 vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration may be suitably formulated to give controlled release of the active compound. For buccal administration the compositions may take the  
30 form of tablets or lozenges formulated in conventional manner. For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane,

dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of e.g., gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as  
5 lactose or starch.

The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or  
10 aqueous vehicles, and may contain formulating agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as cocoa butter or  
15 other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or  
20 hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt. Other suitable delivery systems include microspheres which offer the possibility of local noninvasive delivery of drugs over an extended period of time. This technology utilizes microspheres of precapillary size which can be injected via a coronary catheter into any  
25 selected part of the e.g. heart or other organs without causing inflammation or ischemia. The administered therapeutic is slowly released from these microspheres and taken up by surrounding tissue cells (e.g. endothelial cells).

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be  
30 permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For topical administration, the



oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art. A wash solution can be used locally to treat an injury or inflammation to accelerate healing.

The compositions may, if desired, be presented in a pack or dispenser device which  
 5 may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

#### 4.7. Assays to Identify Therapeutics

Based on the identification of mutations that cause or contribute to the development  
 10 of a disease or disorder that is associated with an IL-1 polymorphism or haplotype, the invention further features cell-based or cell free assays for identifying therapeutics. In one embodiment, a cell expressing an IL-1 receptor, or a receptor for a protein that is encoded by a gene which is in linkage disequilibrium with an IL-1 gene, on the outer surface of its cellular membrane is incubated in the presence of a test compound alone or in the presence  
 15 of a test compound and another protein and the interaction between the test compound and the receptor or between the protein (preferably a tagged protein) and the receptor is detected, e.g., by using a microphysiometer (McConnell et al. (1992) Science 257:1906). An interaction between the receptor and either the test compound or the protein is detected by the microphysiometer as a change in the acidification of the medium. This assay system  
 20 thus provides a means of identifying molecular antagonists which, for example, function by interfering with protein-receptor interactions, as well as molecular agonist which, for example, function by activating a receptor.

Cellular or cell-free assays can also be used to identify compounds which modulate expression of an IL-1 gene or a gene in linkage disequilibrium therewith, modulate  
 25 translation of an mRNA, or which modulate the stability of an mRNA or protein. Accordingly, in one embodiment, a cell which is capable of producing an IL-1, or other protein is incubated with a test compound and the amount of protein produced in the cell medium is measured and compared to that produced from a cell which has not been contacted with the test compound. The specificity of the compound vis a vis the protein can  
 30 be confirmed by various control analysis, e.g., measuring the expression of one or more control genes. In particular, this assay can be used to determine the efficacy of antisense, ribozyme and triplex compounds.

Cell-free assays can also be used to identify compounds which are capable of interacting with a protein, to thereby modify the activity of the protein. Such a compound can, e.g., modify the structure of a protein thereby effecting its ability to bind to a receptor. In a preferred embodiment, cell-free assays for identifying such compounds consist  
5 essentially in a reaction mixture containing a protein and a test compound or a library of test compounds in the presence or absence of a binding partner. A test compound can be, e.g., a derivative of a binding partner, e.g., a biologically inactive target peptide, or a small molecule.

Accordingly, one exemplary screening assay of the present invention includes the  
10 steps of contacting a protein or functional fragment thereof with a test compound or library of test compounds and detecting the formation of complexes. For detection purposes, the molecule can be labeled with a specific marker and the test compound or library of test compounds labeled with a different marker. Interaction of a test compound with a protein or fragment thereof can then be detected by determining the level of the two labels after an  
15 incubation step and a washing step. The presence of two labels after the washing step is indicative of an interaction.

An interaction between molecules can also be identified by using real-time BIA (Biomolecular Interaction Analysis, Pharmacia Biosensor AB) which detects surface plasmon resonance (SPR), an optical phenomenon. Detection depends on changes in the  
20 mass concentration of macromolecules at the biospecific interface, and does not require any labeling of interactants. In one embodiment, a library of test compounds can be immobilized on a sensor surface, e.g., which forms one wall of a micro-flow cell. A solution containing the protein or functional fragment thereof is then flown continuously over the sensor surface. A change in the resonance angle as shown on a signal recording,  
25 indicates that an interaction has occurred. This technique is further described, e.g., in BIA technology Handbook by Pharmacia.

Another exemplary screening assay of the present invention includes the steps of (a) forming a reaction mixture including: (i) an IL-1 or other protein, (ii) an appropriate receptor, and (iii) a test compound; and (b) detecting interaction of the protein and receptor.  
30 A statistically significant change (potentiation or inhibition) in the interaction of the protein and receptor in the presence of the test compound, relative to the interaction in the absence of the test compound, indicates a potential antagonist (inhibitor). The compounds of this assay can be contacted simultaneously. Alternatively, a protein can first be contacted with a

test compound for an appropriate amount of time, following which the receptor is added to the reaction mixture. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison.

5           Complex formation between a protein and receptor may be detected by a variety of techniques. Modulation of the formation of complexes can be quantitated using, for example, detectably labeled proteins such as radiolabeled, fluorescently labeled, or enzymatically labeled proteins or receptors, by immunoassay, or by chromatographic detection.

10           Typically, it will be desirable to immobilize either the protein or the receptor to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of protein and receptor can be accomplished in any vessel suitable for containing the reactants. Examples include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion  
15           protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtitre plates, which are then combined with the receptor, e.g. an <sup>35</sup>S-labeled  
20           receptor, and the test compound, and the mixture incubated under conditions conducive to complex formation, e.g. at physiological conditions for salt and pH, though slightly more stringent conditions may be desired. Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly (e.g. beads placed in scintillant), or in the supernatant after the complexes are subsequently dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by  
25           SDS-PAGE, and the level of protein or receptor found in the bead fraction quantitated from the gel using standard electrophoretic techniques such as described in the appended examples. Other techniques for immobilizing proteins on matrices are also available for use in the subject assay. For instance, either protein or receptor can be immobilized utilizing conjugation of biotin and streptavidin. Transgenic animals can also be made to identify  
30           agonists and antagonists or to confirm the safety and efficacy of a candidate therapeutic. Transgenic animals of the invention can include non-human animals containing a restenosis causative mutation under the control of an appropriate endogenous promoter or under the control of a heterologous promoter.

The transgenic animals can also be animals containing a transgene, such as reporter gene, under the control of an appropriate promoter or fragment thereof. These animals are useful, e.g., for identifying drugs that modulate production of an IL-1 protein, such as by modulating gene expression. Methods for obtaining transgenic non-human animals are well known in the art. In preferred embodiments, the expression of the restenosis causative mutation is restricted to specific subsets of cells, tissues or developmental stages utilizing, for example, cis-acting sequences that control expression in the desired pattern. In the present invention, such mosaic expression of a protein can be essential for many forms of lineage analysis and can additionally provide a means to assess the effects of, for example, expression level which might grossly alter development in small patches of tissue within an otherwise normal embryo. Toward this end, tissue-specific regulatory sequences and conditional regulatory sequences can be used to control expression of the mutation in certain spatial patterns. Moreover, temporal patterns of expression can be provided by, for example, conditional recombination systems or prokaryotic transcriptional regulatory sequences. Genetic techniques, which allow for the expression of a mutation can be regulated via site-specific genetic manipulation in vivo, are known to those skilled in the art.

The transgenic animals of the present invention all include within a plurality of their cells a causative mutation transgene of the present invention, which transgene alters the phenotype of the "host cell". In an illustrative embodiment, either the cre/loxP recombinase system of bacteriophage P1 (Lakso et al. (1992) PNAS 89:6232-6236; Orban et al. (1992) PNAS 89:6861-6865) or the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) Science 251:1351-1355; PCT publication WO 92/15694) can be used to generate in vivo site-specific genetic recombination systems. Cre recombinase catalyzes the site-specific recombination of an intervening target sequence located between loxP sequences. loxP sequences are 34 base pair nucleotide repeat sequences to which the Cre recombinase binds and are required for Cre recombinase mediated genetic recombination. The orientation of loxP sequences determines whether the intervening target sequence is excised or inverted when Cre recombinase is present (Abremski et al. (1984) J Biol. Chem. 259:1509-1514); catalyzing the excision of the target sequence when the loxP sequences are oriented as direct repeats and catalyzes inversion of the target sequence when loxP sequences are oriented as inverted repeats.

Accordingly, genetic recombination of the target sequence is dependent on expression of the Cre recombinase. Expression of the recombinase can be regulated by promoter elements which are subject to regulatory control, e.g., tissue-specific, developmental stage-specific, inducible or repressible by externally added agents. This regulated control will result in genetic recombination of the target sequence only in cells where recombinase expression is mediated by the promoter element. Thus, the activation of expression of the causative mutation transgene can be regulated via control of recombinase expression.

Use of the cre/loxP recombinase system to regulate expression of a causative mutation transgene requires the construction of a transgenic animal containing transgenes encoding both the Cre recombinase and the subject protein. Animals containing both the Cre recombinase and the restenosis causative mutation transgene can be provided through the construction of "double" transgenic animals. A convenient method for providing such animals is to mate two transgenic animals each containing a transgene.

Similar conditional transgenes can be provided using prokaryotic promoter sequences which require prokaryotic proteins to be simultaneously expressed in order to facilitate expression of the transgene. Exemplary promoters and the corresponding transactivating prokaryotic proteins are given in U.S. Pat. No. 4,833,080.

Moreover, expression of the conditional transgenes can be induced by gene therapy-like methods wherein a gene encoding the transactivating protein, e.g. a recombinase or a prokaryotic protein, is delivered to the tissue and caused to be expressed, such as in a cell-type specific manner. By this method, the transgene could remain silent into adulthood until "turned on" by the introduction of the transactivator.

In an exemplary embodiment, the "transgenic non-human animals" of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonal target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonal target cell. The specific line(s) of any animal used to practice this invention are selected for general good health, good embryo yields, good pronuclear visibility in the embryo, and good reproductive fitness. In addition, the haplotype is a significant factor. For example, when transgenic mice are to be produced, strains such as C57BUJ6 or FVB lines are often used (Jackson Laboratory, Bar Harbor, Me.). Preferred strains are those with H-2.sup.b, H-2.sup.d or H-2.sup.q haplotypes such as C57BL/6 or DBA/1. The line(s) used to practice

this invention may themselves be transgenics, and/or may be knockouts (i.e., obtained from animals which have one or more genes partially or completely suppressed).

In one embodiment, the transgene construct is introduced into a single stage embryo. The zygote is the best target for microinjection. In the mouse, the male pronucleus  
5 reaches the size of approximately 20 micrometers in diameter which allows reproducible injection of 1-2 pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster et al. (1985) PNAS 82:4438-4442). As a consequence, all cells of the transgenic animal will carry the incorporated transgene. This  
10 will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene.

Normally, fertilized embryos are incubated in suitable media until the pronuclei appear. At about this time, the nucleotide sequence comprising the transgene is introduced into the female or male pronucleus as described below. In some species such as mice, the  
15 male pronucleus is preferred. It is most preferred that the exogenous genetic material be added to the male DNA complement of the zygote prior to its being processed by the ovum nucleus or the zygote female pronucleus. It is thought that the ovum nucleus or female pronucleus release molecules which affect the male DNA complement, perhaps by replacing the protamines of the male DNA with histones, thereby facilitating the  
20 combination of the female and male DNA complements to form the diploid zygote. Thus, it is preferred that the exogenous genetic material be added to the male complement of DNA or any other complement of DNA prior to its being affected by the female pronucleus. For example, the exogenous genetic material is added to the early male pronucleus, as soon as possible after the formation of the male pronucleus, which is when the male and female  
25 pronuclei are well separated and both are located close to the cell membrane. Alternatively, the exogenous genetic material could be added to the nucleus of the sperm after it has been induced to undergo decondensation. Sperm containing the exogenous genetic material can then be added to the ovum or the decondensed sperm could be added to the ovum with the transgene constructs being added as soon as possible thereafter.

30 Introduction of the transgene nucleotide sequence into the embryo may be accomplished by any means known in the art such as, for example, microinjection, electroporation, or lipofection. Following introduction of the transgene nucleotide sequence into the embryo, the embryo may be incubated in vitro for varying amounts of time, or

reimplanted into the surrogate host, or both. In vitro incubation to maturity is within the scope of this invention. One common method in to incubate the embryos in vitro for about 1-7 days, depending on the species, and then reimplant them into the surrogate host.

For the purposes of this invention a zygote is essentially the formation of a diploid  
5 cell which is capable of developing into a complete organism. Generally, the zygote will be comprised of an egg containing a nucleus formed, either naturally or artificially, by the fusion of two haploid nuclei from a gamete or gametes. Thus, the gamete nuclei must be ones which are naturally compatible, i.e., ones which result in a viable zygote capable of undergoing differentiation and developing into a functioning organism. Generally, a euploid  
10 zygote is preferred. If an aneuploid zygote is obtained, then the number of chromosomes should not vary by more than one with respect to the euploid number of the organism from which either gamete originated.

In addition to similar biological considerations, physical ones also govern the amount (e.g., volume) of exogenous genetic material which can be added to the nucleus of  
15 the zygote or to the genetic material which forms a part of the zygote nucleus. If no genetic material is removed, then the amount of exogenous genetic material which can be added is limited by the amount which will be absorbed without being physically disruptive. Generally, the volume of exogenous genetic material inserted will not exceed about 10 picoliters. The physical effects of addition must not be so great as to physically destroy the  
20 viability of the zygote. The biological limit of the number and variety of DNA sequences will vary depending upon the particular zygote and functions of the exogenous genetic material and will be readily apparent to one skilled in the art, because the genetic material, including the exogenous genetic material, of the resulting zygote must be biologically capable of initiating and maintaining the differentiation and development of the zygote into  
25 a functional organism.

The number of copies of the transgene constructs which are added to the zygote is dependent upon the total amount of exogenous genetic material added and will be the amount which enables the genetic transformation to occur. Theoretically only one copy is required; however, generally, numerous copies are utilized, for example, 1,000-20,000  
30 copies of the transgene construct, in order to insure that one copy is functional. As regards the present invention, there will often be an advantage to having more than one functioning copy of each of the inserted exogenous DNA sequences to enhance the phenotypic expression of the exogenous DNA sequences.

Any technique which allows for the addition of the exogenous genetic material into nucleic genetic material can be utilized so long as it is not destructive to the cell, nuclear membrane or other existing cellular or genetic structures. The exogenous genetic material is preferentially inserted into the nucleic genetic material by microinjection. Microinjection of  
5 cells and cellular structures is known and is used in the art.

Reimplantation is accomplished using standard methods. Usually, the surrogate host is anesthetized, and the embryos are inserted into the oviduct. The number of embryos implanted into a particular host will vary by species, but will usually be comparable to the number of off spring the species naturally produces.

10 Transgenic offspring of the surrogate host may be screened for the presence and/or expression of the transgene by any suitable method. Screening is often accomplished by Southern blot or Northern blot analysis, using a probe that is complementary to at least a portion of the transgene. Western blot analysis using an antibody against the protein encoded by the transgene may be employed as an alternative or additional method for  
15 screening for the presence of the transgene product. Typically, DNA is prepared from tail tissue and analyzed by Southern analysis or PCR for the transgene. Alternatively, the tissues or cells believed to express the transgene at the highest levels are tested for the presence and expression of the transgene using Southern analysis or PCR, although any tissues or cell types may be used for this analysis.

20 Alternative or additional methods for evaluating the presence of the transgene include, without limitation, suitable biochemical assays such as enzyme and/or immunological assays, histological stains for particular marker or enzyme activities, flow cytometric analysis, and the like. Analysis of the blood may also be useful to detect the presence of the transgene product in the blood, as well as to evaluate the effect of the  
25 transgene on the levels of various types of blood cells and other blood constituents.

Progeny of the transgenic animals may be obtained by mating the transgenic animal with a suitable partner, or by in vitro fertilization of eggs and/or sperm obtained from the transgenic animal. Where mating with a partner is to be performed, the partner may or may not be transgenic and/or a knockout; where it is transgenic, it may contain the same or a  
30 different transgene, or both. Alternatively, the partner may be a parental line. Where in vitro fertilization is used, the fertilized embryo may be implanted into a surrogate host or incubated in vitro, or both. Using either method, the progeny may be evaluated for the presence of the transgene using methods described above, or other appropriate methods.



The transgenic animals produced in accordance with the present invention will include exogenous genetic material. Further, in such embodiments the sequence will be attached to a transcriptional control element, e.g., a promoter, which preferably allows the expression of the transgene product in a specific type of cell.

5           Retroviral infection can also be used to introduce the transgene into a non-human animal. The developing non-human embryo can be cultured in vitro to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenich, R. (1976) PNAS 73:1260-1264). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Manipulating the Mouse Embryo, Hogan eds.  
10 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner et al. (1985) PNAS 82:6927-6931; Van der Putten et al. (1985) PNAS 82:6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, supra; Stewart et al. (1987)  
15 EMBO J 6:383-388). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner et al. (1982) Nature 298:623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic non-human animal. Further, the founder may contain various retroviral insertions of the transgene at different positions  
20 in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line by intrauterine retroviral infection of the midgestation embryo (Jahner et al. (1982) supra).

A third type of target cell for transgene introduction is the embryonal stem cell (ES). ES cells are obtained from pre-implantation embryos cultured in vitro and fused with  
25 embryos (Evans et al. (1981) Nature 292:154-156; Bradley et al. (1984) Nature 309:255-258; Gossler et al. (1986) PNAS 83:9065-9069; and Robertson et al. (1986) Nature 322:445-448). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retrovirus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells thereafter  
30 colonize the embryo and contribute to the germ line of the resulting chimeric animal. For review see Jaenisch, R. (1988) Science 240:1468-1474.

The present invention is further illustrated by the following examples which should not be construed as limiting in any way. The contents of all cited references (including

literature references, issued patents, published patent applications as cited throughout this application) are hereby expressly incorporated by reference. The practice of the present invention will employ, unless otherwise indicated, conventional techniques that are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Molecular Cloning A Laboratory Manual, (2nd ed., Sambrook, Fritsch and Maniatis, eds., Cold Spring Harbor Laboratory Press: 1989); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Oligonucleotide Synthesis (M. J. Gait ed., 1984); U.S. Pat. No. 4,683,195; U.S. Pat. No. 4,683,202; and Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds., 1984).

## 5. Examples

The following examples further support, but do not exclusively represent, preferred embodiments of the present invention.

### Example 1IL-1 Gene Locus Mapping and Characterization

The six novel genes encoding proteins with the IL-1 fold have been identified. The classical family are involved in inflammatory signaling. Clone-based and radiation hybrid mapping has placed all six novel genes close to or within the same cluster as the three original gene family members (*IL1A*, *IL1B*, *IL1RN*), in a ~400 kb interval on chromosome 2. We have combined the incomplete public database sequence with our own sequence to generate a reference sequence and map that encompasses all of the novel genes, allowing determination of the gene structures, precise localisation of exons and determination of distances between conventional SNP and microsatellite markers. Gene order from centromere to telomere is *IL1A-IL1B-IL1F7-IL1F9-IL1F6-IL1F8-IL1F5-IL1F10-IL1RN*, of which *IL1A*, *IL1B* and *IL1F8* only are transcribed towards the centromere. The gene order relates to the evolutionary relationship between the genes. Key features of exon boundaries are conserved. There is no evidence for other IL-1 family members within the cluster.

Recently, it was shown that the most closely related receptor to IL-1R1, known as IL-1 receptor related protein 2 (IL-1Rrp2, gene *IL1RL2*) confers responsiveness to IL-1F9 on transfected cells, and that the response is very effectively inhibited by IL-1F5 (IL-1L1), which most closely resembles IL-1ra. The interaction with IL-1F5 seems to be of high affinity. Both IL-1F5 and IL-1F9 are relatively abundant in epithelia, and it has been suggested that they have a role in the regulation of inflammation in this specific compartment. The functions of the other genes are unknown, but low affinity interactions

have been reported between IL-1F7 and IL-18R1 (Pan et al., 2001), and between IL-1F10 and IL-1R1 (Lin et al., 2001). The biological role for the new IL-1 family members is under investigation, but mRNA expression appears to be far more restricted than has been seen in IL-1 $\alpha$ , IL-1 $\beta$ , IL-1ra and IL-18. It is possible, therefore, that the cell types involved in the function of the new IL-1 family members are much more specialised than is the case for IL-1.

### Material and Methods

#### *Sequencing and Sequence Assembly.*

BACs were identified according to partial sequence in the public domain (Lander *et al.*, 2001) as containing *IL1A*, *IL1B*, *IL1RN* and *IL1F5*, which had previously been mapped to a gene cluster (Nicklin et al., 1994; Notwang et al., 1996; Barton et al., 2000). The nine selected BACs were RP11-1I24, RP11-477F18, RP11-554I7, RP11-368A17, RP11-434I13, RP11-67L14, RP11-725J3, RP11-339F22, RP11-97J14 and RP11-65I12. Much of the public data was unfinished and contained no order or orientation information. Aligning the public sequence of individual BACs against one another provided minimal overlap information. In order to generate a minimally tiled scaffold across the region, seven BACs were chosen (RP11-477F18, RP11-554I7, RP11-434I13, RP11-67L14, RP11-725J3, RP11-339F22, RP11-97J14) and sequenced to 3X coverage. Small insert plasmid clones (~3500 bp) were sequenced in both the forward and reverse directions, providing paired reads across clones. PHRED and PHRAP (Ewing et al., 1998; Ewing and Green, 1998) were used for the base calls and assemblies of the seven BACs. Internal contig viewing tools were used to analyze the resulting assemblies. We ordered contigs by matching sequenced contig ends whose paired reads fell on other contig ends. At this low coverage, the BACs assembled to a large number of contigs, but the order and orientation were established. Public data for the seven internally sequenced BACs as well as two externally finished BACs (RP11-1I24 and RP11-65I12) were imported from Genbank. Various software tools were used to compare and align the internal, public, and overlapping sequences, providing order and orientation information across all available data. Contigs were then chosen from these alignments to create as much contiguous sequence as possible across the region and assembled using Sequencher (version 4.0.5).

*Sequence alignment and exon assignment.*

Primer and cDNA sequences were initially matched to genomic sequences with a 2-sequence BLAST routine (Altschul et al., 1997) running on the NCBI server. Exon alignments were made with the est2genome routine (Mott, 1997), running on the HGMP server (Cambridge, UK). The program was set to identify consensus exon boundaries. 5' exons which could not be identified because of their shortness were localised manually to the closest corresponding sequence terminating at a consensus splice donor dinucleotide (GT). No attempt was made to map the 3' ends of non-coding regions as mRNA size data are largely not available.

*Results**The Sequence of the IL-1 Cluster.*

A 900 kilobase region was assembled into 14 ordered contiguous sequences combining the internal and public sequences. The telomeric portion of this sequence contains the gene *PAX8*. Subsequently, a shorter region, composed of seven of the contigs, totalling 496 kb was extracted from the region. Recent updates of the public database have allowed us to patch five of the six gaps in the sequence. (see Fig 17). We have submitted an annotated sequence, as described in this report, of 495475 nucleotides to the public databases (accession \*\*\*). The single remaining gap (marked "gap" on Fig. 17) is centromeric of the IL-1 cluster. The sequence is not of finished quality but provides a framework for the finished sequence and allows us to examine the structure of the genes within the IL-1 cluster. The new map is consistent with previously published maps (Nicklin et al., 1994; Nothwang et al., 1996) but differs substantially from the incomplete public genome assembly project (Lander et al., 2001).

The closest identified flanking gene towards the centromere is unrelated to IL-1. It is the plasma membrane phosphate transporter *SLC20A1* (previously identified as the human homologue of the gibbon-ape leukaemia virus receptor, *GLVR1*, accession XM\_002217), which maps between 63 kb and 45 kb to the left of the origin on Fig. 1. Towards the telomere of the cluster lies *TIC* (Accession NM\_012455), which is most probably an ARF6-selective guanine nucleotide exchange factor (MN and Tomas Klenka, manuscript in preparation), at the telomeric flank. Its map position is shown in Fig. 17.

### Gene Structures

We have mapped all of the IL-1 family cDNA sequences onto the genomic sequence (Fig. 17) where the extent of the genes is shown with black rectangles. Figure 17 shows a map of the IL-1 gene cluster. Scale bars (in kb) are provided above and below the data to aid alignments. The sources of the data described are indicated by the top three lines. “Novel sequence” was determined entirely at Genome Therapeutics. “Public DB” indicates sequence taken from Genbank. “Combined sequence” were assembled from a combination of the two sources. Above the bar representing the contig, the positions of previously described polymorphic markers (summarised by Cox *et al.*, 1996 and di Giovine *et al.*, 2000) are indicated with labelled arrows. The single unfilled gap is also indicated. CpG-rich regions as defined in the text are indicated “CpGr”. The probable sites of the rare cutter restriction enzyme sites clusters that were used in previous mapping are also marked as “Xrec”, “Yrec?”, and “Zrec”. The extent of the mapping of the cDNA sequences of Figure 18 onto the contig is indicated by the solid black rectangles below the contig line, except the non-cytokine gene *TIC*, which is marked grey. The positions of the coding sequences for CE1, CE2 and CE3 are indicated by vertical bars. The gene symbols are followed or preceded by a chevron to indicate the direction of transcription. Figure 17 further shows the detailed structure of the IL-1 Cluster. Each gene is listed in order from centromere to telomere. “Gene” the conventional locus name for the gene. “Orientation” is either “forward”, where the deposited sequence is the sense strand, or “reverse” where it is the anti-sense. “Position” is the nucleotide numbers on the deposited sequence corresponding to each exon.

When cDNA sequences are known to be incomplete, likely extensions of exons are marked with “<” and “>” symbols. “Exon” is the name we are assigning to each exon, based on its presence in the cDNA for one of the corresponding transcripts; thus *IL1RN-a4/b5/c6* is the 4<sup>th</sup> exon of cDNA a (X52015), the 5<sup>th</sup> of cDNA b (M55646) and the 6<sup>th</sup> of cDNA c. The identities of the corresponding mRNAs has been agreed (\*\*). An asterisk (\*) against adjacent entries indicates that two exons share a splice donor site as a result of the use of alternative promoters. “Exon Boundaries” are the 15 nucleotide sequences within the exon that flank the intron. An ellipsis (...) at either end indicates that the exon is likely to be incomplete because the cDNA sequence has been truncated.. “Exon type” indicates the coding potential of the exon: 5’N, 5’-non-translated region; 5’S0, potentially translated 5’ short open reading frame; Ps, peptide presequence ( indicates that this has been proposed);

cs, unconserved coding sequence; CE, conserved exons; 3'N, 3'-non translated region. An ellipsis indicates that the exon assignment is probably not complete and that some or all non-coding sequence has been omitted. "Coding" indicates the amino-acid sequence encoded by each exon. The exons are identified by the cDNA name and accession  
5 (indicated at the top of the box) that they compose. The coding capacity of each exon is indicated in lower case. Italicised residues are encoded partly on the next exon. The numeric superscript indicates the number of bases in the stated exon contained within the codon. The residue is omitted from the next exon. The nucleotides in the bridging codon are indicated by italics in the "exon boundaries" box. Where the succeeding exon is  
10 alternative, the bridging residue may change. This is indicated in parenthesis. Underscored residues are from the terminal complete codons of the exon and their codons are underscored in the "exon boundaries" box. An asterisk indicates translational termination.

*IL1A* is the most centromeric gene and is transcribed towards the centromere, as is the  
15 adjacent gene, *IL1B*. The remaining genes, ending with *IL1RN*, the most telomeric member of the cluster, are transcribed towards the telomere, with the exception of *IL1F8*. The three last exons of each gene, which we have called common exons (CE)1, 2 and 3, encode the IL-1-homologous domain (as shown in Figure 18 and defined elsewhere) and fall in compact regions within the sequence. CE1, CE2 and CE3 are indicated by vertical bars in  
20 Fig. 1, but at the resolution of Fig. 17, some cannot be distinguished. Additional exons with little or no coding content extend the span of most of the genes considerably. The largest spans are *IL1RN* and *IL1F8*. In the latter case, the first non-coding exon is 20 kb telomeric of the rest of the gene. Details of the mapping of the genes are given in , along with the encoded peptide sequence from each exon. Where splice variants exist, this information  
25 allows the reader to assemble the different possible protein forms. It is currently uncertain whether all of these forms are likely to be biologically relevant (see Discussion).

Figure 18 (sheets 1-7) shows the alignment of the encoded sequence of the three common exons of the ten known members of the IL-1 family. In each case the common exons are the last three of a transcript; e.g. exons 5, 6 and 7 out of the 7 exons of IL-1 $\alpha$ .  
30 Alignment was done by eye by seeking amino acid identities and blocks of similar residues. Gaps were then minimised. Crystallographic data for IL-1 $\beta$  and IL-1ra were incorporated and used further to refine the alignment. Translations of the three common exon portions are shown in order. Numbers indicate the first and last codons of the mature product that

are encoded by each exon. Gene products are listed in accordance with their probable phylogeny. (!) indicates that processing at a proteolytic site yields the mature protein, but that some of the presequence is also encoded within the first common exon. Blocked residues are common to at least three sequences. For simplicity, similarity is not indicated.

5 For IL-1 $\beta$  and IL-1ra, on the first line below the coding sequence, (labelled “crystallography”) the approximate positions of the ends of the  $\beta$ -sheets are indicated by vertical bars and the span of the sheet is shaded grey and labelled with the number of the sheet. In the next line (labelled “contacts”), numbers indicate the domain of the IL-1R interacting with the side chain of each residue. A numbered residue contains at least one  
10 heavy atom (C, N, O, S) that lies within 4 Å of a heavy atom of the type I IL-1 receptor (PDB data), as visualized with the program RasMol (Sayle and Milner-White). In the line below IL-1F5 (labelled NMR), a (^) indicates residues of IL-1F5 that show a strong (>0.7 ppm) upfield shift in their  $\alpha$ -<sup>13</sup>C NMR signal, which is taken to indicate a high probability of its residing within a  $\beta$ -sheet. The final line of the block (labelled “consensus”) indicates,  
15 in lower case, residues that occur at least 7/10 times in that position. Where capitalized, the residue is present in all cases. An ellipsis indicates that sheet 1 of a particular sequence probably begins on a previous exon. (\*) indicates translational termination.

#### *CpG-rich regions.*

20 The program CpGplot (Larsen *et al.*, 1992) was used to identify five potential CpG islands with  $\geq 60\%$  C+G content,  $\geq 60\%$  of the expected frequency of the CpG dinucleotide and of  $\geq 300$  nucleotides in length. With the exceptions of the first and the two last CpG-rich sequences, these regions are short and probably do not constitute “CpG islands”. There are thus no CpG islands in the IL-1 cluster. We have attempted to locate the clusters of  
25 restriction sites that were used previously for physical mapping (Nicklin *et al.*, 1994). CpG-rich sequences are labeled *CpGr* in Fig. 1. Two are further labeled *Xrec* and *Zrec*. These two regions contain the specific rare cutter restriction sites that were identified previously, and so probably correspond to the cluster’s flanks, as previously assigned. The sequence data gives a length of 392 kb compared with the previous estimate of 430 kb from  
30 Southern hybridization of restriction digests of genomic DNA. A close pairing of *Nae* I and *Eag* I sites, which was previously used to map *IL1B* is seen around the site labelled *Yrec?*, but was not selected by the program *CpGplot*, even with less stringent parameters. Only *Xrec* and *Zrec* mark substantial CpG islands. Database searching and the public genome

annotation effort has not yet revealed genes to be associated with either of these loci. One possibility is that *Zrec* marks an unrecognised upstream exon of TIC, a non-cytokine gene that is abundantly expressed in all tissues tested (Tomas Klenka and MN, unpublished data).

5

*Polymorphic markers in the IL-1 Cluster.*

We have placed the polymorphisms in this region that have been described previously (indicated by arrows in Fig. 17 and listed in Figure 19). This has allowed us to reassess disequilibrium data described previously (Cox *et al.*, 1998). Our analysis gives a slightly better correlation coefficient between map distance and decay of disequilibrium (data not shown).

10

*Scanning the IL-1 Cluster for further IL-1-like genes.*

We investigated whether there are further IL-1-like sequences within the IL-1 cluster. Because of its relatively small size, the genomic sequence of the cluster was amenable to very low stringency searching with the BLAST algorithm (Altschul *et al.*, 1997). The NCBI server for two-sequence BLAST comparisons was used with its default settings, except that the sensitivity was raised to expect 5000 hits per genome (from its default value of 10). Translations of individual exons were submitted for TBLASTN analysis of the IL-1 cluster genomic sequence. This algorithm performs a search of the coding sequence against the six possible reading frames derived from the genomic sequence fragment. We assumed that exon structure would be conserved, so matches were subsequently discounted if they were interrupted with stop codons.

15

20

Because it is one of the more distantly related sequences, we searched first with the CE3 of *IL1A*. This matched only itself. CE3 of *IL1B* returned CE3 of all known family members on the IL-1 cluster except *IL1A*. One uninterrupted hit was found, but it shared only 6 identical putative residues, was longer than typical for a CE3 and actually lay in reverse orientation within *IL1B*. The sequence was discounted as there was no evidence for a corresponding potential upstream CE2. We next searched with CE3 of *IL1F5*, which also returned all of the CE3s except *IL1A*. One long, potential CpG-rich exon lacked the conserved core residues of CE3. As another outlier, we used CE3 of *IL18* (accession XM\_041373). This returned *IL1F5* from the IL-1 cluster and no novel sequences. We next tested CE2 (exons 6) from *IL1A* and *IL1B*. The former returned only itself, the latter

30



returned *IL1F6*, *IL1F8*, *IL1F9* and *IL1F10* and no other sequence. CE2 of *IL1F5* returned *IL1RN*, *IL1F6*, *IL1F9* and *IL1F10*, but no novel uninterrupted exons. CE2 of *IL18* returned none. Finally CE2 of *IL1F9* was tested. It returned CE2 of *IL1F6*, *IL1F8*, *IL1RN*, *ILF10* and no other sequence. We conclude that there are no further IL-1 family genes within the  
 5 IL-1 cluster unless they have either a highly divergent sequence or differ from all of the other family members in having a more fragmented exon structure.

### *Evolutionary Considerations*

To investigate the phylogeny of the IL-1 family, We ran the program Tree-Puzzle  
 10 (Strimmer and von Haesler, 1996) on the alignment of CE3 shown in Figure 2a. IL-18 was set as the outgroup member of the family. The result was visualised in a radial dendrogram (Page, 1996) shown in Figure 3.

### Example 2: Case -Cohort Study of Inflammatory Genes and Coronary Heart Disease

15 (a sub-study of the Atherosclerosis Risk in Communities (ARIC) project)

ARIC is a prospective cohort study designed to investigate the etiology and natural history of atherosclerosis, the etiology of clinical atherosclerotic diseases, and variation in cardiovascular risk factors, medical care, and disease by race, gender, place and time.

The ARIC cohort consists of a probability sample of 15,792 individuals, age 45-64  
 20 years at baseline, from four U.S. communities. ILGN has approval to genotype all participants in the ARIC program as appropriate to meet the objectives of the two collaborative sub-studies. In our ongoing study of incident cardiovascular events we now have DNA samples from 955 ARIC participants who have experienced acute clinical events along with a randomly sampled cohort control group. These samples represent all incident  
 25 cardiovascular cases during the first 11 years of longitudinal monitoring. The genotyping of all samples was recently completed and partial results are available. These results demonstrate significant associations between risk of clinical events and IL-1( +4845) allele 2 for subjects with total cholesterol (TC) <200 mg/dl. Key aspects of these findings include:

- 30 • +4845 genotype significantly associated with clinical events (Survival Analysis Relative Risk ~ 4.0, p<.01)
- Analysis included all ages

- In a multivariate model, the IL-1 genotype findings were independent of age, gender, smoking, race, diabetes, hypertension, BMI, LDL, HDL
- Number of subjects included with TC<200 was 955

Locus: IL1A (+4845), total cholesterol <200mg/dl stratum

5 *Time to first acute coronary artery disease event*

Within each table there are three models. The first is the crude model which has just the genotype variables. This is identified by “Crude” in the Adjustment column. The models with “Group 1” in the Adjustment column adjust for age, sex and race/center. Those with “Group 2” in the Adjustment column adjust for age, sex, race/center, current smoker

10 (yes/no), diabetic (yes/no), hypertensive (yes/no), LDL cholesterol, and HDL cholesterol.

Comparing ‘1.2’ and ‘2.2’ against the baseline of ‘1.1’

Adjustment	Genotype	BETA	SE	T	P	RR	LOWER_95	UPPER_95
Crude	1.2	-0.27372	0.21851	-1.25269	0.21032	0.76054	0.49559	1.16714
	2.2	0.34378	0.40769	0.84324	0.39909	1.41027	0.63426	3.13573
Group 1	1.2	-0.05382	0.24299	-0.22151	0.82469	0.94760	0.58856	1.52566
	2.2	0.72707	0.47274	1.53798	0.12405	2.06901	0.81913	5.22602
Group 2	1.2	-0.02903	0.28294	-0.10261	0.91827	0.97139	0.55789	1.6913
	2.2	1.38022	0.49320	2.79850	0.00513	3.97577	1.51217	10.4530

15

Comparing ‘2.2’ against the baseline of ‘1.1’ and ‘1.2’ together

Adjustment	Genotype	BETA	SE	T	P	RR	LOWER_95	UPPER_95
Crude	2.2	0.46712	0.39696	1.17673	0.23930	1.59539	0.73277	3.47352
Group 1	2.2	0.75024	0.45891	1.63485	0.10208	2.11752	0.86139	5.20541
Group 2	2.2	1.39344	0.47638	2.92505	0.00344	4.02869	1.58365	10.2487

20 Comparing ‘2.2’ against the baseline of ‘1.1’ with subjects having ‘1.2’ excluded

Adjustment	Genotype	BETA	SE	T	P	RR	LOWER_95	UPPER_95
Crude	2.2	0.34181	0.40774	0.83831	0.40186	1.40750	0.63295	3.12987
Group 1	2.2	0.75725	0.48656	1.55634	0.11963	2.13241	0.82168	5.53400
Group 2	2.2	1.60027	0.57185	2.79840	0.00514	4.95436	1.61517	15.1970

Example 3 The San Francisco Study of Osteoporotic Fractures (SOF)

The Multi-center Study of Osteoporotic Fractures under the direction of Dr. Steven Cummings at the University of California in San Francisco, consists of a large cohort of women of European/Caucasian origin from 4 different clinical centers. These women have been examined since 1986 for various medical and lifestyle findings, including hip, wrist and spine fractures and changes in bone mineral density in lumbar spine and femoral neck. At baseline visit (1986/1987) all participants (n = 9,704) were 65 year or older, ambulatory and not institutionalized. Blood samples were collected from approximately 4,000 subjects and stored at -70°C for DNA analysis.

A recent analysis of cause of death in the SOF cohort determined that IL-1A(4845) allele 2 was significantly associated with early death from cardiovascular disease.

		Relative Risk	LOWER CI	UPPER CI	UNIT	PVALUE
CVD death	N=452;					
IL1A_1	IL-1A 1.2 VS 1.1	1.03	0.49	2.167	1	0.937
IL1A_2	IL-1A 2.2 VS 1.1	3.138	1.203	8.184	1	0.0194
RAGE2	ADJUSTED CURRENT AGE	2.431	1.842	3.209	5	0

Example 4 Functional Analysis of the +4845 IL-1 SNPs

+4845 SNP is a non-synonymous SNP (i.e. a naturally-occurring polymorphism which alters the amino acid of and leads to an amino acid change in the IL-1a cytokine). The variant proteins are expressed in insect cells using baculoviral vectors and analyzed for structural and functional differences. The variant cDNAs used for the expression of the protein in insect cells and in mammalian cells are confirmed by sequence analysis to only contain one SNP leading to an amino acid change. Here are 2 pieces of data related to this SNP.

In the Western Blot analysis (see Figure 12), we provide data to show that the 2 variants of the IL-1a cytokine are processed differently with calpain digestion. Calpain is an enzyme known to cleave the full length IL-1a cytokine (31 kDa) to form the mature protein (17 kDa). The allele-1 (Ala) IL-1a cytokine gives rise to a single 17 kDa molecule, whereas, the allele-2 (Ser) IL-1a cytokine yields 2 bands, one which is identical in size to

the band found with the allele-1 but additionally, it also gives rise to another band which is slightly larger in molecular weight. This result indicates that there is a structural difference in the 2 variants. We also postulate that the Ala to Ser mutation leads to differential post-translational modification of the proteins, for example, differences in phosphorylation or myristolation. This amino acid change could lead to an alteration (addition or removal) of the recognition signal for the post-translation modification.

Fibroblast cells stably transfected with the ala and ser variant cDNAs in expression vectors were found to have a different rate of proliferation. The allele-2 variant has a faster growth rate than the allele-1 variant that supports our claim that allele-2 is predictive of a proinflammatory profile. (see Figure 13). Accordingly, the altered amino acid in the allele-2 variant shows evidence of a more potent proinflammatory cytokine than the allele-1 variant.

Example 5 Systematic Functional Analysis of the IL-1A, IL-1B and IL-1RN SNPs

In this example, selected IL-1A, IL-1B, and IL-1RN polymorphisms are constructed in a background of otherwise “wild type” IL-1 sequence and the effects are measured in a fibroblast cell line.

Transcriptional analysis of IL-1A, IL-1B, IL-1RN gene promoter SNPs by reporter-promoter constructs. Each gene’s data is in separate figure (i.e. Figures 14, 15 and 16 respectively). The Figures 14, 15, and 16 panel A (and Figure 15 D) shows the SNPs and the various allele-2 mutations that were created in separate luciferase constructs and also the different lengths of promoter- luciferase constructs annotated with the SNPs investigated in the transfection analysis. In addition, we also provide luciferase assay results for only the functional SNPs that show an altered activity of the gene transcription with respect to the wild type (allele-1 at all loci). For the B gene, we also provide data for the functional SNPs in a backbone where SNP#14 (-511) and SNP#2 (-31) are also allele-2.

Note that these constructs were tested in a fibroblast cell line (i.e. WI38- which models a specific role of IL-1 in the inflammatory response. Accordingly, other cell lines which model other mechanistic aspects of IL-1-mediated inflammatory diseases and disorders will be specifically tested. For example, human cell monocyte cell lines (e.g. U937) and human keratinocyte cell line (e.g. A143) and in a human osteoblast cell line (e.g. to investigate affects upon osteoporosis IL-1 inflammatory processes.

Example 6 Annotation of the IL-1 Gene Cluster SNPs

We have further annotated polymorphisms throughout the IL-1 gene cluster (see Figures 8-11). As these polymorphisms occur within established IL-1 haplotypes as herein supported (see Figures 1-7), they provide compositions and methods which are supported in the instant application.

5

#### **Equivalents and Incorporation by Reference**

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific polypeptides, nucleic acids, methods, assays and reagents described herein. Such equivalents are considered to be within the scope of this invention and are covered by the following claims.

10

The instant application includes numerous citations to learned texts, published articles and patent applications as well as issued U.S. and foreign patents. The entire contents of all of these citations are hereby incorporated by reference herein.

15